



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 140719

TO: James Schultz
Location: REM-2D18/2C18
Art Unit: 1635
Wednesday, December 22, 2004

Case Serial Number: 10/001851

From: Edward Hart
Location: Biotech-Chem Library
REM-1A55
Phone: 571-272-2512

edward.hart@uspto.gov

Search Notes

Examiner Schultz,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

no page blank (uspto)

STIC-Biotech/ChemLib

140719

From: Schultz, James
Sent: Sunday, December 19, 2004 9:56 PM
To: STIC-Biotech/ChemLib
Subject: Seq search 10/001,851

Hello,
Could you please search SEQ ID NO 2 in the amino acid databases (603 aa long, no interference search needed) in the above entitled application?
Thanks
Doug Schultz

James Douglas Schultz, PhD
AU 1635 (Biotechnology)
Patent Examiner
United States Patent and Trademark Office
(Office) REM 2D18
(Mail) REM 2C18
(571) 272-0763

RECEIVED
DEC 20 2005
STIC-Biotech Division
(STIC)

STAFF USE ONLY

Searcher: _____
Searcher Phone: 2- _____
Date Searcher Picked up: 12/20/04
Date Completed: 12/20/04
Searcher Prep/Rev. Time: _____
Online Time: _____

Type of Search
NA Sequence: # _____
AA Sequence: # 1
Structure: # _____
Bibliographic: _____
Litigation: _____
Patent Family: _____
Other: _____

Vendors and cost where applicable
STN: _____
DIALOG: _____
QUESTEL/ORBIT: _____
LEXIS/NEXIS: _____
SEQUENCE SYSTEM: _____
WWW/Internet: _____
Other(Specify): _____

Image Blank (uspto)

Pending Nucleic Acid and Pending Amino Acid database searches generate two sets of results each. The Pending databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches. Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions .rnpm and .rnpn. Searches run against the Amino Acid Pending database produce two sets of results, with the extensions .rapm and .rapn.

Because they contain data that is confidential, the results of Pending database searches should not be left in the case .

Page Blank (uspto)

Protein Sequence Searches - 10/8/04

All of the sequence databases on the ABSS have been updated. A change has occurred in the protein databases.

- Two protein databases, SPTREMBL and SwissProt, are now produced as a single, merged database called UniProt.
- Results from UniProt have the file extension **.rup**.
- Sequences in UniProt are identified by the same ID that had been used in SPTREMBL or SwissProt.
- In instances where the database curators have determined that an SPTREMBL record and a SwissProt record represent the same sequence, the two records have been merged into one. Both IDs are present in the record. Any differences found between the two sequences are recorded in the FT (feature table) fields.

If you have any questions regarding these changes or your results, please contact any STIC searcher.

nis Page Blank (uspto)



STIC SEARCH RESULTS FEEDBACK FORM

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher* or contact:

Mary Hale, Information Branch Supervisor
571-272-2507 Remsen E01 D86

Voluntary Results Feedback Form

➤ I am an examiner in Workgroup: Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library Remsen Bldg.



This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2004, 13:55:08 ; Search time 159 Seconds

(Without alignments)
1360.464 Million cell updates/sec

Title: US-10-001-851-2

Perfect score: 3278
Sequence: 1 MRKREKRLQAVLVLAALV.....TQGMLEPHNNTYLEKFNEN 603

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_23Sep04:*

1: geneseqp1980s:.*
2: geneseqp1980s:.*
3: geneseqp2000s:.*
4: geneseqp2001s:.*
5: geneseqp2002s:.*
6: geneseqp2003as:.*
7: geneseqp2003bs:.*
8: geneseqp2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3278	100.0	603	4	AAU07778	AAU07778 Human nov
2	3278	100.0	603	5	AAU99098	AAU99098 Human gly
3	3272	99.8	603	7	ADCC3587	ADCC3587 Human N-a
4	3242.5	98.9	631	4	AAU07777	AAU07777 Human nov
5	2771	84.5	506	4	AAU07772	AAU07772 Human nov
6	2771	84.5	506	6	AAU07781	AAU07781 Carboxydr
7	2746.5	83.8	535	4	AAU07771	AAU07771 Human nov
8	1893	57.7	339	4	AAU94733	AAU94733 Human pro
9	1877	57.3	366	4	AAU07776	AAU07776 Human nov
10	1818	55.5	407	4	AAU62600	AAU62600 Human ace
11	1403.5	42.8	319	4	ABG08187	ABG08187 Novel hum
12	1396	42.6	275	7	ADCE25865	ADCE25865 Human N-a
13	1392	42.5	276	8	ADOL17613	ADOL17613 Human sof
14	1385.5	42.3	335	4	ABG08189	ABG08189 Novel hum
15	1370	41.8	269	4	AAU07770	AAU07770 Human nov
16	1360	41.5	666	4	ABH66873	ABH66873 Drosophi1
17	1360	41.5	666	4	ABH58867	ABH58867 Drosophi1
18	1311	40.0	321	4	AAU07774	AAU07774 Human nov
19	1186	35.2	1252	4	ABH59363	ABH59363 Drosophi1
20	1155	35.2	499	4	ABG08190	ABG08190 Novel hum
21	1125	34.3	559	7	ADP65304	ADP65304 Human pol
22	1125	34.3	561	3	ABH43561	ABH43561 Human can
23	1116	34.0	559	7	ADB79836	ADB79836 Rat polyp
24	1116	34.0	559	7	ADB79796	ADB79796 Rat polyp
25	1116	34.0	559	7	ADH62526	ADH62526 Rat Prote

26	1115	34.0	517	2	AAH66402	AAH66402 GalNAc-tr
27	1115	34.0	517	2	AAH6489	AAH6489 Honeybee
28	1115	34.0	559	2	AAH6397	AAH6397 Cattle Ga
29	1115	34.0	559	2	AAH6401	AAH6401 GalNAc-tr
30	1115	34.0	559	2	AAH6484	AAH6484 Bovine N-
31	1113.5	34.0	630	4	ABH64271	ABH64271 Drosophi1
32	1112	33.9	571	7	ADCL0116	ADCL0116 Human NOV
33	1106	33.7	556	5	ABG32379	ABG32379 Novel hum
34	1106	33.7	556	7	ADH25918	ADH25918 Novel N-a
35	1106	33.7	557	7	ADH76895	ADH76895 Human N-a
36	1096.5	33.5	555	7	ADCL0114	ADCL0114 Human NOV
37	1085.5	33.1	573	8	ADP04927	ADP04927 Sea squit
38	1050	32.0	657	4	AAH88419	AAH88419 Human mem
39	1048	32.0	478	6	ABH41518	ABH41518 Human DIT
40	1048	32.0	478	7	ADL2696	ADL2696 Human DIT
41	1045.5	31.9	658	4	AAH78711	AAH78711 Human pro
42	1040	31.7	930	7	ADH55070	ADH55070 Rat Prote
43	1035.5	31.6	591	4	ABH62319	ABH62319 Drosophi1
44	1026	31.3	578	6	ABH07527	ABH07527 Human N-a
45	1024.5	31.3	188	7	ADH25974	ADH25974 N-acetyl1g

ALIGNMENTS

RESULT 1	AAU07778	standard; protein; 603 AA.
ID	AAU07778	
AC	AAU07778	
DT	04-DEC-2001	(first entry)
DS	Human novel transferase protein, NHP #21.	
XX	Human; transferase; breast cancer; prostate cancer; immunogen;	
KW	gene therapy; antilease.	
OS	Homo sapiens.	
XX	WO200164903-A2.	
PN	07-SEP-2001.	
PD	28-FEB-2001; 2001WO-US006460.	
XX	29-FEB-2000; 2000US-0185920P.	
PR	02-MAR-2000; 2000US-0186558P.	
PR	24-MAR-2000; 2000US-0191849P.	
XX	(TEXT-) LEXICON GENETICS INC.	
FA	Donoho G, Hilbun E, Turner CA, Friedrich G, Abulin A;	
XX	Zambrwitz B, Sands AT, Walke DW, Wilgnowski NL, Hu Y, Kieke JA;	
PI	Potter DG;	
XX	WPI, 2001-550185/61.	
DR	N-PSDB; AAS12625.	
XX	Novel nucleic acid sequences encoding novel human proteins useful for	
PT	diagnosis, drug screening; clinical trial monitoring and treatment of	
PT	diseases and disorders.	
PT	Claim 5; Page 56-58; 60pp; English.	
PS	The invention relates to isolated nucleic acids encoding novel human	
XX	transferase proteins (NHP). The nucleic acids and proteins are useful for	
CC	diagnosis, drug screening, clinical trial monitoring and treatment of	
CC	diseases and disorders e.g. breast and prostate cancer. NHPs can also be	
CC	useful for augmenting the efficacy of chemotherapeutic agents used in	
CC	treatment of breast or prostate cancer. The nucleic acid is also useful	
CC	in NHP gene regulation, and as antisense primers in amplification	
CC	reactions of NHP gene sequences. NHPs are useful for producing	

CC antibodies. The present sequence represents a novel human transferase
CC which has sequence similarity to N-acetyl-galactosaminyltransferase
XX
SQ Sequence 603 AA;

Query Match 100.0%; Score 3278; DB 4; Length 603;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 603; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MRREKRLIQAVLVLAALVLPVNGLMALYEROPDPTGGSGAIVAPAAAGGSHSRK 60
DB 1 MRREKRLIQAVLVLAALVLPVNGLMALYEROPDPTGGSGAIVAPAAAGGSHSRK 60
QY 61 KTFPLDGGQKLKDMHDKAIRDACRVNGEGQRPYPMIDARVDAVRENGENIVYSDK 120
DB 61 KTFPLDGGQKLKDMHDKAIRDACRVNGEGQRPYPMIDARVDAVRENGENIVYSDK 120
QY 121 ISLNSRLPDIRHPNCNCKRYLETLPNTSIIIPHNCGSSLLRTVHSLNRSPPELVAEI 180
DB 121 ISLNSRLPDIRHPNCNCKRYLETLPNTSIIIPHNCGSSLLRTVHSLNRSPPELVAEI 180
QY 181 VLVDDSDREHLKKPLEDMALFPSSVRIITKKREGIIRRMIGASVATGDTVITFLDSHC 240
DB 181 VLVDDSDREHLKKPLEDMALFPSSVRIITKKREGIIRRMIGASVATGDTVITFLDSHC 240
QY 241 EAVNMVLPPLLDRIARNRKTIVCPMIDVIDHDDFRYETQAGDAMRGAFDWMYTKRIPIP 300
DB 241 EAVNMVLPPLLDRIARNRKTIVCPMIDVIDHDDFRYETQAGDAMRGAFDWMYTKRIPIP 300
QY 301 PELQKADPSDPFESSPVWAGGLPAVDRKFMWELGSDYDGLIWDGEOYEISFKVMCGGRM 360
DB 301 PELQKADPSDPFESSPVWAGGLPAVDRKFMWELGSDYDGLIWDGEOYEISFKVMCGGRM 360
QY 361 EDIPGCRVGHVYKRYVPYKIPAGVSLARNLKVAAEVMMDYATYIYQREYRHLAAGDV 420
DB 361 EDIPGCRVGHVYKRYVPYKIPAGVSLARNLKVAAEVMMDYATYIYQREYRHLAAGDV 420
QY 421 AVOKLRSSLNCSFKPMFTKIAMDLPKFPVPEPPAAAGELIRNVGTGCACTKIGALG 480
DB 421 AVOKLRSSLNCSFKPMFTKIAMDLPKFPVPEPPAAAGELIRNVGTGCACTKIGALG 480
QY 481 SPRLTSGCVARGGEAAAMNNQVFTFTWRREDIRGDDPQHTKKCFDASHSTSPVLYDCHS 540
DB 481 SPRLTSGCVARGGEAAAMNNQVFTFTWRREDIRGDDPQHTKKCFDASHSTSPVLYDCHS 540
QY 541 MKGNOLMKYRKDITLVHPVSGCMDCSESDHRIFMNTCNSSLTQOMLFHTNSTYLEKF 600
DB 541 MKGNOLMKYRKDITLVHPVSGCMDCSESDHRIFMNTCNSSLTQOMLFHTNSTYLEKF 600
QY 601 NRN 603
DB 601 NRN 603
```

RESULT 2
AAU99098
AAU99098 standard; protein; 603 AA.

AAU99098;
30-AUG-2002 (first entry)
Human glycosyl transferase 47169.

Human; glycosyl transferase; 47169; 33935; cancer; carbohydrate storage;
diabetes mellitus; hypoglycaemia; arthritis; rheumatism;
autoimmune disorder; systemic lupus erythematosus; Grave's disease;
myasthenia gravis; insulin resistance; scleroderma; rheumatoid arthritis;
autoimmune infertility; tumourigenesis.

Homo sapiens.
OS
XX
XX
PN WO200240657-A2.

XX 23-MAY-2002.
XX
XX 20-NOV-2001; 2001WO-US047575.
XX
XX 20-NOV-2000; 2000US-0249939P.
XX
XX 20-NOV-2000; 2000US-0249939P.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Meyers R, Williamson M;
XX
XX WPI; 2002-508326/54.
XX
XX N-PSDB; ABK86093.

Use of modulator of activity of novel glycosyl transferase proteins,
47169/33935 proteins, for making a medicament for modulating ability of
cell to affect glycosylation state of lipid or polypeptide target in
cell.
Claim 27; Fig 1; 153pp; English.

The invention relates to the use of a modulator of the activity of an
enzyme such as 47169 or 33935 protein (a novel glycosyl transferase
protein) for making a medicament for modulating the ability of a cell to
affect the glycosylation state of a lipid target or polypeptide target in
a cell. Also included is a method of assessing (M1) if a test compound is
useful for modulating at least one phenomenon (P) such as non-covalent
binding between a protein and one of a cell, a virus and another protein;
cell signaling; cell differentiation; tumourigenesis; cell adhesion; cell
motility; cell-to-cell interaction; cell invasivity; cell proliferation;
gene transcription, and an immune response, comprising: (a) adding the
test compound to a first composition comprising a 603 residue 47169
polypeptide sequence (S2), or a sequence at least 90 % identical to a 492
residue 33935 polypeptide sequence (S12) both given in the specification
; and (b) comparing the activity in the first composition and in a second
composition that is substantially identical to the first composition,
except that it lacks the test compound, whereby a difference in the
activity in the first and second compositions is an indication that the
test compound is useful for modulating the phenomenon. The method is
useful for making a medicament for modulating the ability of a cell (e.g.
human endothelial cell such as lung cell, breast cell or colon cell,
preferably a tumour cell) to affect the glycosylation state of a target
such as lipid or polypeptide. The identified test compound is useful for
treating disorders such as diabetes mellitus, hypoglycaemia, arthritis,
rheumatism, autoimmune disorders (e.g. systemic lupus erythematosus,
Grave's disease, myasthenia gravis, insulin resistance, rheumatoid
arthritis, scleroderma and autoimmune infertility), tumourigenesis,
cancer and tumour metastasis. The present sequence represents glycosyl
transferase 47169

Sequence 603 AA;

Query Match 100.0%; Score 3278; DB 5; Length 603;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 603; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MRREKRLIQAVLVLAALVLPVNGLMALYEROPDPTGGSGAIVAPAAAGGSHSRK 60
DB 1 MRREKRLIQAVLVLAALVLPVNGLMALYEROPDPTGGSGAIVAPAAAGGSHSRK 60
QY 61 KTFPLDGGQKLKDMHDKAIRDACRVNGEGQRPYPMIDARVDAVRENGENIVYSDK 120
DB 61 KTFPLDGGQKLKDMHDKAIRDACRVNGEGQRPYPMIDARVDAVRENGENIVYSDK 120
QY 121 ISLNSRLPDIRHPNCNCKRYLETLPNTSIIIPHNCGSSLLRTVHSLNRSPPELVAEI 180
DB 121 ISLNSRLPDIRHPNCNCKRYLETLPNTSIIIPHNCGSSLLRTVHSLNRSPPELVAEI 180
QY 181 VLVDDSDREHLKKPLEDMALFPSSVRIITKKREGIIRRMIGASVATGDTVITFLDSHC 240
DB 181 VLVDDSDREHLKKPLEDMALFPSSVRIITKKREGIIRRMIGASVATGDTVITFLDSHC 240
QY 241 EAVNMVLPPLLDRIARNRKTIVCPMIDVIDHDDFRYETQAGDAMRGAFDWMYTKRIPIP 300
```



```

Db      241 EAVVNMPLPLDRIANRKTIVCPMIDVIDHDDFRYETQAGDAMRGAFDWMYKRIPIR 300
Qy      301 PELQKADPSDPFSPVWAGGLFVDRKMFWEELGYPDGLIWCGEYEISFKYMMCCGRM 360
Db      301 PELQKADPSDPFSPVWAGGLFVDRKMFWEELGYPDGLIWCGEYEISFKYMMCCGRM 360
Qy      361 EDIPGSRVGHVYKRYVPYKVPAGVSLARNLKRVAEVMDEYAEYIYORREYHLSAGDV 420
Db      361 EDIPGSRVGHVYKRYVPYKVPAGVSLARNLKRVAEVMDEYAEYIYORREYHLSAGDV 420
Qy      421 AVQKLRSSLNCKSPFKMFTKIAMDLPEKYPVPEPPAAAMGEIRNVGTGLCADTKIGALG 480
Db      421 AVQKLRSSLNCKSPFKMFTKIAMDLPEKYPVPEPPAAAMGEIRNVGTGLCADTKIGALG 480
Qy      481 SPLRLGCVRGGEAAMNNMNVFTFTWRREDIRPGDPQHTKKFCFDAISHTSPVTLVDCHS 540
Db      481 SPLRLGCVRGGEAAMNNMNVFTFTWRREDIRPGDPQHTKKFCFDAISHTSPVTLVDCHS 540
Qy      541 MKGNQLMKRYKDKTILYHPVSGSCMDCSESDHRIFMNTCNPSLSLTQWLFPHNTSVLEKF 600
Db      541 MKGNQLMKRYKDKTILYHPVSGSCMDCSESDHRIFMNTCNPSLSLTQWLFPHNTSVLEKF 600
Qy      601 NRN 603
Db      601 NRN 603

RESULT 3
ADC35387
ID ADC35387 standard; protein; 603 AA.
XX
XX ADC35387;
XX
XX 18-DEC-2003 (first entry)
XX
DE Human N-acetyl-galactosamine transferase (GalNAc) T13 protein.
XX
XX N-acetyl-galactosamine transferase; GalNAc; alpha1-bond,
XX genetic engineering; GalNAc-T13; cyostatic; cancer treatment;
XX cancer diagnosis; gene therapy; human; enzyme.
XX
XX Homo sapiens.
XX
XX MO2003057887-A1.
XX
XX 17-JUL-2003.
XX
XX 06-JAN-2003; 2003WO-JP000008.
XX
XX 28-DEC-2001; 2001JP-00401507.
XX 05-JUN-2002; 2002JP-00163832.
XX 12-JUL-2002; 2002JP-00203696.
XX
XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
XX (AMSH) AMERSHAM BIOSCIENCES KK.
XX (FURE) FUJIREBIO INC.
XX
XX Narimatsu H, Zhang Y, Gotoh M;
XX
XX WPI; 2003-587133/55.
XX DR N-PSDB; ADC35394.
XX
XX Novel UDP-N-acetyl-D-galactosamine:polypeptide N-acetyl-galactosamine
XX transferase and encoded nucleic acid, applicable in identifying O-bonded
XX sugar-attached proteins, gene diagnosis, and cancer therapy.
XX
XX Claim 1; SEQ ID NO 3; 153pp; Japanese.
XX
XX This invention relates to a novel protein with an activity of
XX transferring N-acetyl-galactosamine (GalNAc) via an alpha1-bond to a
XX hydroxyl group of serine or threonine in a protein or a peptide sequence.
XX The invention provides gene and protein sequences for these enzymes to

```

```

CC enable genetic engineering or production of these enzymes. Seven genes
CC and their proteins are described by the invention, GalNAc-T11 to 17. The
CC enzymes of the invention may have cyostatic activity, and hence may be
CC used for treatment or diagnosis of cancer, and the sequences described
CC may be useful in gene therapy. The present sequence is the partial
CC sequence of the human GalNAc-T13 transferase protein of the invention.
XX
XX Sequence 603 AA:
Qy      1 MRKREKRLQAVLALVTLVPLNVLMLYRROPDGRPGSGAVALPAAAGGSHSRK 60
Db      1 MRKREKRLQAVLALVTLVPLNVLMLYRROPDGRPGSGAVALPAAAGGSHSRK 60
Qy      61 KTFPLDGGOKLQDMHDKA1RRDAQVNGEGQRPYPMTDABRVDAQYRENGFNIVSDK 120
Db      61 KTFPLDGGOKLQDMHDKA1RRDAQVNGEGQRPYPMTDABRVDAQYRENGFNIVSDK 120
Qy      121 ISLNRSLPDIRHPNCKSKYLETLPMNTSIIIPPHNKGWSLRTVHSLNRSPELVAEI 180
Db      121 ISLNRSLPDIRHPNCKSKYLETLPMNTSIIIPPHNKGWSLRTVHSLNRSPELVAEI 180
Qy      181 VLVDDFSDEHLLKKPLLEDVYALFSPVRILLRTKKRBLIRRMGASVATGDTPLDSHC 240
Db      181 VLVDDFSDEHLLKKPLLEDVYALFSPVRILLRTKKRBLIRRMGASVATGDTPLDSHC 240
Qy      241 EAVVNMPLPLDRIANRKTIVCPMIDVIDHDDFRYETQAGDAMRGAFDWMYKRIPIR 300
Db      241 EAVVNMPLPLDRIANRKTIVCPMIDVIDHDDFRYETQAGDAMRGAFDWMYKRIPIR 300
Qy      301 PELQKADPSDPFSPVWAGGLFVDRKMFWEELGYPDGLIWCGEYEISFKYMMCCGRM 360
Db      301 PELQKADPSDPFSPVWAGGLFVDRKMFWEELGYPDGLIWCGEYEISFKYMMCCGRM 360
Qy      361 EDIPGSRVGHVYKRYVPYKVPAGVSLARNLKRVAEVMDEYAEYIYORREYHLSAGDV 420
Db      361 EDIPGSRVGHVYKRYVPYKVPAGVSLARNLKRVAEVMDEYAEYIYORREYHLSAGDV 420
Qy      421 AVQKLRSSLNCKSPFKMFTKIAMDLPEKYPVPEPPAAAMGEIRNVGTGLCADTKIGALG 480
Db      421 AVQKLRSSLNCKSPFKMFTKIAMDLPEKYPVPEPPAAAMGEIRNVGTGLCADTKIGALG 480
Qy      481 SPLRLGCVRGGEAAMNNMNVFTFTWRREDIRPGDPQHTKKFCFDAISHTSPVTLVDCHS 540
Db      481 SPLRLGCVRGGEAAMNNMNVFTFTWRREDIRPGDPQHTKKFCFDAISHTSPVTLVDCHS 540
Qy      541 MKGNQLMKRYKDKTILYHPVSGSCMDCSESDHRIFMNTCNPSLSLTQWLFPHNTSVLEKF 600
Db      541 MKGNQLMKRYKDKTILYHPVSGSCMDCSESDHRIFMNTCNPSLSLTQWLFPHNTSVLEKF 600
Qy      601 NRN 603
Db      601 NRN 603

RESULT 4
AAU07777
ID AAU07777 standard; protein; 631 AA.
XX
XX AAU07777;
XX
XX 04-DEC-2001 (first entry)
XX
DE Human novel transferase protein, NHP #20.
XX
XX Human; transferase; breast cancer; prostate cancer; immunogen;
XX gene therapy; antisense.
XX
XX Homo sapiens.
XX

```

FH Key Location/Qualifiers
 FT MISC-difference 630. .631
 FT /note="Encoded by AATGGAAC"
 XX
 XX WO200164903-A2.
 XX
 PD 07-SEP-2001.
 XX
 PF 28-FEB-2001; 2001WO-US006460.
 XX
 PR 29-FEB-2000; 2000US-0185920P.
 PR 02-MAR-2000; 2000US-0186558P.
 PR 24-MAR-2000; 2000US-0191849P.
 PA (LEXI-) LEXICON GENETICS INC.
 XX
 PI Donoho G, Hilbun E, Turner CA, Friedrich G, Abulin A;
 PI Zambrowicz B, Sands AT, Walke DW, Wilganowski NL, Hu Y, Kieke JA;
 PI Potter DG;
 XX
 DR MPI; 2001-550185/61.
 DR N-PSDB; AAS12624.
 XX
 PT Novel nucleic acid sequences encoding novel human proteins useful for
 PT diagnosis, drug screening, clinical trial monitoring and treatment of
 PT diseases and disorders.
 XX
 PS Claim 6; Page 54-55; 60pp; English.
 XX
 CC The invention relates to isolated nucleic acids encoding novel human
 CC transferase proteins (NHP). The nucleic acids and proteins are useful for
 CC diagnosis, drug screening, clinical trial monitoring and treatment of
 CC diseases and disorders e.g. breast and prostate cancer. NHPs can also be
 CC useful for augmenting the efficacy of chemotherapeutic agents used in
 CC treatment of breast or prostate cancer. The nucleic acid is also useful
 CC in NHP gene regulation, and as antisense primers in amplification
 CC reactions of NHP gene sequences. NHPs are useful for producing
 CC antibodies. The present sequence represents a novel human transferase
 CC which has sequence similarity to N-acetyl-galactosaminyltransferase
 CC
 SO Sequence 631 AA;
 Query Match 98.9%; Score 3242.5; DB 4; Length 631;
 Best Local Similarity 95.4%; Pred. No. 0;
 Matches 601; Conservative 0; Mismatches 0; Indels 29; Gaps 1;
 QY 1 MRKREKLLQAVLVLAALVLLPVGIMALYRROPDGTGGGAAVAPAGGSGSHSRK 60
 DB 1 MRKREKLLQAVLVLAALVLLPVGIMALYRROPDGTGGGAAVAPAGGSGSHSRK 60
 QY 61 KTFPFLDGGOKLXDMHDKXEARIRDAORVNGEGORPYPMTDABRVDAVRENGFNIVSDK 120
 DB 61 KTFPFLDGGOKLXDMHDKXEARIRDAORVNGEGORPYPMTDABRVDAVRENGFNIVSDK 120
 QY 121 ISLNRSLPDIRHNCNCKRYLETLPNTSIIIPHNESMSLRTVHSVLRSPPELVAEI 180
 DB 121 ISLNRSLPDIRHNCNCKRYLETLPNTSIIIPHNESMSLRTVHSVLRSPPELVAEI 180
 QY 181 VLVDVDSDBREHLKKPLIEDYVALTPSVRIILTKKREGIRIRRMGASVATDVTTFDSDHC 240
 DB 181 VLVDVDSDBREHLKKPLIEDYVALTPSVRIILTKKREGIRIRRMGASVATDVTTFDSDHC 240
 QY 241 EAVVNMVLPPLLDRIARNRKTIIVCPMTDIVDHDPRYETOGADMRAFGPMEMVYKXIPIP 300
 DB 241 EAVVNMVLPPLLDRIARNRKTIIVCPMTDIVDHDPRYETOGADMRAFGPMEMVYKXIPIP 300
 QY 301 PELQKADPSDPFESPVWAGGLFAVDRKMFWELOGYDGLIHWGEOYEISFK----- 352
 DB 301 PELQKADPSDPFESPVWAGGLFAVDRKMFWELOGYDGLIHWGEOYEISFKGHLMLPRL 360
 QY 353 -----VMMCGRMEDIPCSRVGHIYRKYVPYKVPAGVSLARNLK 391
 DB 361 VSNWSDQAVFLPRAFMNLALQVMMCGRMEDIPCSRVGHIYRKYVPYKVPAGVSLARNLK 420

QY 392 RVAEVMWDEYAEIYQRRPEYHLSAGDVAVOKLRSSLNCKSPKFMFKIAMDLPKFYP 451
 DB 421 RVAEVMWDEYAEIYQRRPEYHLSAGDVAVOKLRSSLNCKSPKFMFKIAMDLPKFYP 480
 QY 452 PVPEPAAWGEIRNVGTGLCADTKGALGSPRLBSCVGRGBAAMNNQVFTTWRBDI 511
 DB 481 PVPEPAAWGEIRNVGTGLCADTKGALGSPRLBSCVGRGBAAMNNQVFTTWRBDI 540
 QY 512 RGGDPQHTKKFCFPAISHSPVTLVDCHSMKGNOLMKYRKDTLVHPVSGSCMDGSESDH 571
 DB 541 RGGDPQHTKKFCFPAISHSPVTLVDCHSMKGNOLMKYRKDTLVHPVSGSCMDGSESDH 600
 QY 572 RIFMNTCNPSLSLQOWLFPHNTSVLEKEN 601
 DB 601 RIFMNTCNPSLSLQOWLFPHNTSVLEKEN 630
 RESULT 5
 AAU07772
 ID AAU07772 standard; protein; 506 AA.
 XX
 AC AAU07772;
 XX
 DT 04-DEC-2001 (first entry)
 XX
 DE Human novel transferase protein, NHP #15.
 XX
 KM Human; transferase; breast cancer; prostate cancer; immunogen;
 KM gene therapy; antisense.
 OS Homo sapiens.
 XX
 PN WO200164903-A2.
 XX
 PD 07-SEP-2001.
 XX
 PF 28-FEB-2001; 2001WO-US006460.
 XX
 PR 29-FEB-2000; 2000US-0185920P.
 PR 02-MAR-2000; 2000US-0186558P.
 PR 24-MAR-2000; 2000US-0191849P.
 XX
 PA (LEXI-) LEXICON GENETICS INC.
 XX
 PI Donoho G, Hilbun E, Turner CA, Friedrich G, Abulin A;
 PI Zambrowicz B, Sands AT, Walke DW, Wilganowski NL, Hu Y, Kieke JA;
 PI Potter DG;
 XX
 DR MPI; 2001-550185/61.
 DR N-PSDB; AAS12619.
 XX
 PT Novel nucleic acid sequences encoding novel human proteins useful for
 PT diagnosis, drug screening, clinical trial monitoring and treatment of
 PT diseases and disorders.
 XX
 PS Claim 8; Page 47-48; 60pp; English.
 XX
 CC The invention relates to isolated nucleic acids encoding novel human
 CC transferase proteins (NHP). The nucleic acids and proteins are useful for
 CC diagnosis, drug screening, clinical trial monitoring and treatment of
 CC diseases and disorders e.g. breast and prostate cancer. NHPs can also be
 CC useful for augmenting the efficacy of chemotherapeutic agents used in
 CC treatment of breast or prostate cancer. The nucleic acid is also useful
 CC in NHP gene regulation, and as antisense primers in amplification
 CC reactions of NHP gene sequences. NHPs are useful for producing
 CC antibodies. The present sequence represents a novel human transferase
 CC which has sequence similarity to N-acetyl-galactosaminyltransferase
 CC
 SO Sequence 506 AA;
 Query Match 84.5%; Score 2771; DB 4; Length 506;
 Best Local Similarity 100.0%; Pred. No. 8.2e-274;

	Matches	506:	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Oy		98	MTDAERVDQAVRENGFNIIYVSDKISLNSSLPDIRHPNCSKRYYLTETLENTSIIIPFHNEG		157					
Dd		1	MTDAERVDQAVRENGFNIIYVSDKISLNSSLPDIRHPNCSKRYYLTETLENTSIIIPFHNEG		60					
Oy		158	WSSILRTTHSVLANSPEELVAEIVLVDDPSREHLKKPELYMALPESVRILTRTKRGL		217					
Dd		61	WSSLRLRTTHSVLANSPEELVAEIVLVDDPSREHLKKPELYMALPESVRILTRTKRGL		120					
Oy		218	IRTEMLGASVATGDVITFLDSHCENANMVLPELDRIARNRKTIYCPMIIDVIDHDDPFYE		277					
Dd		121	IRTEMLGASVATGDVITFLDSHCENANMVLPELDRIARNRKTIYCPMIIDVIDHDDPFYE		180					
Oy		278	TOAGDAMRGAFDWEMYRYKRIPIPELOKADPSDPESPVMAGLPFAVDKKFWELGYDP		337					
Dd		181	TOAGDAMRGAFDWEMYRYKRIPIPELOKADPSDPESPVMAGLPFAVDKKFWELGYDP		240					
Oy		338	GLETWGSGQYIEISFYVMCGGRMEI PCSRHGIYRKVPYPKPVPGUSLAARLKRAVEYW		397					
Dd		241	GLETWGSGQYIEISFYVMCGGRMEI PCSRHGIYRKVPYPKPVPAVASIAENLRVAEVW		300					
Oy		398	MDEYAELYORRPPEXRNHSAGDVAAQQKLRSLSLNCKSKFMPTKTAMPDKPYPPVEBPA		457					
Dd		301	MDEYAELYORRPPEXRNHSAGDVAAQQKLRSLSLNCKSKFMPTKTAMPDKPYPPVEBPA		360					
Oy		458	AAMGEIRNVGTGLCADTRKGALGPSPLREGCVRGGEAANNMMQYFTFTWRBDIRPGDPQ		517					
Dd		361	AAMGEIRNVGTGLCADTRKGALGPSPLREGCVRGGEAANNMMQYFTFTWRBDIRPGDPQ		420					
Oy		518	HTKKFCFPALIHSTSPVTLYDCHSMKGNOLMKYRKDKTLYHPVSGSCMDCSSDRIFMNT		577					
Dd		421	HTKKFCFPALIHSTSPVTLYDCHSMKGNOLMKYRKDKTLYHPVSGSCMDCSSDRIFMNT		480					
Oy		578	CNPSSLTQOQLFEHNTSNVYLEKFNN 603							
Dd		481	CNPSSLTQOQLFEHNTSNVYLEKFNN 506							
RESULT 6										
AAG79781	ID	AAG79781	standard; protein; 506 AA.							
XX	AA	AAG79781;								
DT	16-APR-2003	(first entry)								
XX		Carbohydrate-associated protein (CHOP) -3.								
DE		Human; carbohydrate associated polypeptides; CHOP; cancer;								
KM		carbohydrate metabolism; diabetes; anaemia; hypoglycaemia; obesity;								
KM		glycogen storage disease; neurological dysfunction; gene therapy;								
KM		cell proliferation; actinic keratosis; arteriosclerosis; inflammation;								
KM		atherosclerosis; cirrhosis; hepatitis; psoriasis; autoimmune disorder;								
KM		acquired immunodeficiency syndrome; AIDS; humanized; transgenic;								
KM		Addison's disease; allergy; asthma; contact dermatitis; bronchitis;								
KM		Hashimoto's thyroiditis; infection; reproduction; endometriosis;								
KM		polycystic ovary syndrome; ovarian hyperstimulation syndrome;								
KM		Down syndrome; cystic fibrosis; sickle cell anaemia; thalassemia;								
KM		myocarditis; cardiomyopathy; Alzheimer's disease; Parkinson's disease;								
KM		dementia; depression; epilepsy; Tourette's disorder; schizophrenia;								
KM		central nervous system; cerebral palsy; mood; anxiety; knockin.								
XX		Homo sapiens.								
OS										
XX										
FH	Key	Location/Qualifiers								
FT	Peptide	3..503								
FT	/label=	Acetylgalectosaminyl transferase								
FT	/note=	"Identified by BLAST_DOMO"								
FT		8..493								
FT	Peptide	/label=	Acetylgalactosaminyl transferase							
FT	/note=	"Identified by BLAST_DOMO"								
FT		11..492								
FT	Peptide									

FT	/label= Acetylglucosaminyl transferase
FT	/note= "identified by BLAST_DOMO"
FT	51..236
FT	/label= glycosyl transferase sequence
FT	/note= "identified by HMMER_PPRAM"
FT	202..362
FT	/label= N-acetylgalactosaminyl transferase
FT	/note= "identified by BLAST_PRODROM"
FT	364..402
FT	/label= QXM lectin repeat
FT	/note= "identified by HMMER_PPRAM"
FT	417..455
FT	/label= QXM lectin repeat
FT	/note= "identified by HMMER_PPRAM"
FT	456..495
FT	/label= QXM lectin repeat
FT	/note= "identified by HMMER_PPRAM"
XX	
PM	WO200297060-A2.
PD	
PD	05-DEC-2002.
PE	22-MAY-2002; 2002WO-US018354.
PR	
PR	25-MAY-2001; 2001US-0293768P.
PR	01-AUG-2001; 2001US-0309548P.
PR	23-AUG-2001; 2001US-0314400P.
PR	19-OCT-2001; 2001US-0343706P.
PR	07-DEC-2001; 2001US-0337999P.
PA	(INCYT-) INCYTE GENOMICS INC.
P1	Swarnakar A, Gorvad AS, Hafalia AJA, Duggan BM, Emerling BM;
P1	Ison CH, Nguyen DB, Lee EA, Yue H, Forsythe LJ, Li JK,
P1	Thangavelu K, Walla NK, Burford N, Mason PM, Lal PG, Lee S,
P1	Becha SD, Tang YT;
DR	WPI; 2003-140462/13.
DR	N-PSDB; ABAB0833.
PT	
PT	Novel human carbohydate associated polypeptide, useful in diagnosis,
PT	treatment and prevention of carbohydrate metabolism, cell proliferative,
PT	autoimmune/inflammatory, reproductive, and neurological disorders.
PS	
PS	Claim 1; Page 126-27; 141pp; English.
XX	
CC	The sequences given in AAC79779-88 represent human carbohydrate
CC	associated polypeptides (CHOP). The CHOP polypeptides and the nucleotide
CC	sequences encoding them, are useful for diagnosing, treating and
CC	preventing cancer, carbohydrate metabolic disorders (e.g. diabetes,
CC	anemia, hypoglycaemia, obesity, glycogen storage disease, neurological
CC	dysfunctions), cell proliferative disorders (e.g. actinic keratosis,
CC	artherosclerosis, atherosclerosis, cirrhosis, hepatitis, psoriasis),
CC	autoimmune/inflammatory disorders (such as acquired immunodeficiency
CC	syndrome (AIDS), Addison's disease, allergies, asthma, contact
CC	dermatitis, bronchitis, gout, Hashimoto's thyroiditis), viral, bacterial,
CC	fungal, parasitic, protozoal and helminthic infections, reproductive
CC	disorders (e.g. endometriosis, polygenic ovary syndrome, ovarian
CC	hyperstimulation syndrome), genetic disorders (e.g. Down syndrome, cystic
CC	fibrosis, sickle cell anaemia, thalassemia), cardiac disorders (e.g.
CC	myocarditis, cardiomyopathy), neurological disorders (e.g. Alzheimer's
CC	disease, Parkinson's disease, dementia, depression, epilepsy, Tourette's
CC	disorder, schizophrenia), and developmental disorders of central nervous
CC	system (e.g. cerebral palsy), mental disorders (e.g. mood, anxiety). CHOP
CC	proteins are useful in a number of drug screening techniques, and to
CC	analyse the proteome of a tissue or cell type. CHOP cDNA is useful for
CC	creating "knockin" humanized animals or transgenic animals to model human
CC	diseases, in somatic or germ-line gene therapy, to generate a transcript
CC	image of a tissue or cell type, for detecting differences in the
CC	chromosomal location due to translocation, inversion, etc., among normal,
CC	carrier or affected individuals, and as hybridization probes for mapping
XX	naturally occurring genomic sequences

SQ Sequence 506 AA;
 Query Match 84.5%; Score 2771; DB 6; Length 506;
 Best Local Similarity 100.0%; Pred. No. 8.2e-274;
 Matches 506; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 98 MTDARVDAQYRENGFNIVYSDKISLNRSLPDIRHPNCNKRYLETLPNTSIIIPPHNEG 157
 1 MTDARVDAQYRENGFNIVYSDKISLNRSLPDIRHPNCNKRYLETLPNTSIIIPPHNEG 60
 DB 158 WSSLRTVHSLVNRSPPELVAEIVLVDDFSDREHLKKPLEDYMALPFSVAILRTKKREG 217
 61 WSSLRTVHSLVNRSPPELVAEIVLVDDFSDREHLKKPLEDYMALPFSVAILRTKKREG 120
 QY 218 IRTRMGASVATGDTVITFLDSHCENAVNMLPPLLDRIARNRKTIVCPMIDVIDHDDFRYE 277
 121 IRTRMGASVATGDTVITFLDSHCENAVNMLPPLLDRIARNRKTIVCPMIDVIDHDDFRYE 180
 DB 278 TQAGDAMRGAFDWEYMYKRIPIPELOKADPSDFESPVMAGLFPAVDKMFWELEGYDP 337
 181 TQAGDAMRGAFDWEYMYKRIPIPELOKADPSDFESPVMAGLFPAVDKMFWELEGYDP 240
 QY 338 GLEIMGGEQYEISFKVMCGRMEDIPCSRGHYRKYVPAGVSLARNLKRYAEVW 397
 241 GLEIMGGEQYEISFKVMCGRMEDIPCSRGHYRKYVPAGVSLARNLKRYAEVW 300
 DB 398 MDEYAEITYORREYRHLASGDAVAVQKKLSLNCSEFKMFMTKIANDLPKYPVPEPPA 457
 301 MDEYAEITYORREYRHLASGDAVAVQKKLSLNCSEFKMFMTKIANDLPKYPVPEPPA 360
 QY 458 AANGEIRNVGTGLCADTKHGALGSPRLBECVGRGEAANNMNOVFTTWRREDIRPGDPQ 517
 361 AANGEIRNVGTGLCADTKHGALGSPRLBECVGRGEAANNMNOVFTTWRREDIRPGDPQ 420
 DB 518 HTKKFCFDASHSPVLYDCHSKMGKQMLKRYRKDKTYHPVSGSCMDCESDHRLFMNT 577
 421 HTKKFCFDASHSPVLYDCHSKMGKQMLKRYRKDKTYHPVSGSCMDCESDHRLFMNT 480
 QY 578 CNPSLSLTOQWLFEHTNSTVLEKFNRN 603
 481 CNPSLSLTOQWLFEHTNSTVLEKFNRN 506
 DB
 RESULT 7
 AAU07771 standard; protein; 535 AA.
 ID AAU07771;
 AC AAU07771;
 XX
 DT 04-DEC-2001 (first entry)
 DE Human novel transferase protein, NHP #14.
 DB Human novel transferase protein, NHP #14.
 KW Human; transferase; breast cancer; prostate cancer; immunogen;
 KM gene therapy; antilease.
 OS Homo sapiens.
 XX
 PN MO200164903-A2.
 XX
 PD 07-SEP-2001.
 XX
 PF 28-FEB-2001; 2001MO-US006460.
 XX
 PR 29-FEB-2000; 2000US-0185920P.
 PR 02-MAR-2000; 2000US-0186558P.
 PR 24-MAR-2000; 2000US-0191849P.
 XX
 PA (LEXI-) LEXICON GENETICS INC.
 XX
 PI Donoho G, Hlibun E, Turner CA, Friedrich G, Abulin A,
 PI Zambrowicz B, Sands AT, Walke DW, Wilganowski NL, Hu Y, Kiecke JA,
 PI Potter DG;

XX
 DR WPI; 2001-550185/61.
 DR N-PSDB; AAS12618.
 XX
 PT Novel nucleic acid sequences encoding novel human proteins useful for
 PT diagnosis, drug screening, clinical trial monitoring and treatment of
 PT diseases and disorders.
 XX
 PS Claim 7; Page 45-46; 60pp; English.
 CC The invention relates to isolated nucleic acids encoding novel human
 CC transferase proteins (NHP). The nucleic acids and proteins are useful for
 CC diagnosis, drug screening, clinical trial monitoring and treatment of
 CC diseases and disorders e.g. breast and prostate cancer. NHPs can also be
 CC useful for augmenting the efficacy of chemotherapeutic agents used in
 CC treatment of breast or prostate cancer. The nucleic acid is also useful
 CC in NHP gene regulation, and as antisense primers in amplification
 CC reactions of NHP gene sequences. NHPs are useful for producing
 CC antibodies. The present sequence represents a novel human transferase
 CC which has sequence similarity to N-acetyl-galactosaminyltransferase
 CC
 SQ Sequence 535 AA;
 Query Match 83.8%; Score 2746.5; DB 4; Length 535;
 Best Local Similarity 94.6%; Pred. No. 2.9e-271;
 Matches 506; Conservative 0; Mismatches 0; Indels 29; Gaps 1;
 QY 98 MTDARVDAQYRENGFNIVYSDKISLNRSLPDIRHPNCNKRYLETLPNTSIIIPPHNEG 157
 1 MTDARVDAQYRENGFNIVYSDKISLNRSLPDIRHPNCNKRYLETLPNTSIIIPPHNEG 60
 DB 158 WSSLRTVHSLVNRSPPELVAEIVLVDDFSDREHLKKPLEDYMALPFSVAILRTKKREG 217
 61 WSSLRTVHSLVNRSPPELVAEIVLVDDFSDREHLKKPLEDYMALPFSVAILRTKKREG 120
 QY 218 IRTRMGASVATGDTVITFLDSHCENAVNMLPPLLDRIARNRKTIVCPMIDVIDHDDFRYE 277
 121 IRTRMGASVATGDTVITFLDSHCENAVNMLPPLLDRIARNRKTIVCPMIDVIDHDDFRYE 180
 DB 278 TQAGDAMRGAFDWEYMYKRIPIPELOKADPSDFESPVMAGLFPAVDKMFWELEGYDP 337
 181 TQAGDAMRGAFDWEYMYKRIPIPELOKADPSDFESPVMAGLFPAVDKMFWELEGYDP 240
 QY 338 GLEIMGGEQYEISFK-----VMMCGRMEDIPCSR 368
 241 GLEIMGGEQYEISFKGLHMLPRLVSNMPOAVFLPRAPNMLAQVMMCGRMEDIPCSR 300
 DB 369 GHYRKYVPYKVPAGVSLARNLKRYAEVWMDYAEITYORREYRHLASGDAVAVQKKLS 428
 301 GHYRKYVPYKVPAGVSLARNLKRYAEVWMDYAEITYORREYRHLASGDAVAVQKKLS 360
 QY 429 SLNCKSEFKMFMTKIANDLPKYPVPEPPAAANGEIRNVGTGLCADTKHGALGSPRLBEC 488
 361 SLNCKSEFKMFMTKIANDLPKYPVPEPPAAANGEIRNVGTGLCADTKHGALGSPRLBEC 420
 DB 489 VRRGGEAANNMNOVFTTWRREDIRPGDPQHTKKFCFDAISHSPVLYDCHSKMGKQMLK 548
 421 VRRGGEAANNMNOVFTTWRREDIRPGDPQHTKKFCFDAISHSPVLYDCHSKMGKQMLK 480
 QY 549 YRKDKTYHPVSGSCMDCESDHRLFMNTCNPSLSLTOQWLFEHTNSTVLEKFNRN 603
 481 YRKDKTYHPVSGSCMDCESDHRLFMNTCNPSLSLTOQWLFEHTNSTVLEKFNRN 535
 DB
 RESULT 8
 AAB94733 standard; protein; 339 AA.
 ID AAB94733;
 XX
 AC AAB94733;
 XX
 DT 26-JUN-2001 (first entry)
 DE Human protein sequence SEQ ID NO:15766.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
KW Homo sapiens.
OS
XX EP1074617-A2.
PN
XX 07-FEB-2001.
PD
XX 28-JUL-2000; 2000EP-00116126.
PF
XX 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
XX (HELI-) HELIX RES INST.
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
DR WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
XX Claim 8; SEQ ID NO 15766; 2537bp + Sequence listing; English.
XX
XX The present invention describes primer sets for synthesising 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
XX Sequence 339 AA;
SQ
Query Match 57.7%; Score 1893; DB 4; Length 339;
Best Local Similarity 99.7%; Pred. No. 3.1e-184;
Matches 338; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 265 MIDVIDHDDPRYEQADMGAFDMWYVYRIRIPPELOKADSDPESPVMAGGFAY 324
DB 1 MIDVIDHDDPRYEQADMGAFDMWYVYRIRIPPELOKADSDPESPVMAGGFAY 60
QY 325 DRKFMELGCGYDPLGELTWGEGQYEIFKVMWCGRMEDIPCSRGHYTKYVPYKVGAV 384
DB 61 DRKFMELGCGYDPLGELTWGEGQYEIFKVMWCGRMEDIPCSRGHYTKYVPYKVGAV 120
QY 385 SLAANLKRVAEVMWDEVAEYIYQRRPEYRHLASGDVAVQKTLRSSLNCKSFKMFMTKIAY 444
DB 121 SLAANLKRVAEVMWDEVAEYIYQRRPEYRHLASGDVAVQKTLRSSLNCKSFKMFMTKIAY 180
QY 445 DLRFYFVPEVPPAALAKGEIRNVGTGLCADTKHGLSGPLRLGCGVRGEGALMNMNQVFT 504

DB 181 DLRFYFVPEVPPAALAKGEIRNVGTGLCADTKHGLSGPLRLGCGVRGEGALMNMNQVFT 240
QY 505 FTWRREDIRPQDPQHTKKFCFDALSHTSPTLLDCHSMKGNQMLKRYKDKTLVHPVSGSCM 564
DB 241 FTWRREDIRPQDPQHTKKFCFDALSHTSPTLLDCHSMKGNQMLKRYKDKTLVHPVSGSCM 300
QY 565 DCSGSDHRIFMNTCNPSLTQOMLPHTNSTVLEKFN 603
DB 301 DCSGSDHRIFMNTCNPSLTQOMLPHTNSTVLEKFN 339
RESULT 9
AAU07776
ID AAU07776 standard; protein; 366 AA.
XX
XX AAU07776;
AC
XX
XX 04-DEC-2001 (first entry)
DT
XX
XX Human novel transferase protein, NHP #19.
DE
XX
XX Human; transferase; breast cancer; prostate cancer; immunogen;
KW gene therapy; antisense.
XX
XX Homo sapiens.
OS
XX WO200164903-A2.
PN
XX
XX 07-SEP-2001.
PD
XX
XX 28-FEB-2001; 2001WO-US006460.
PF
XX
XX 29-FEB-2000; 2000US-0185320P.
PR 02-MAR-2000; 2000US-0186558P.
PR 24-MAR-2000; 2000US-0191849P.
XX
XX (LEXI-) LEXICON GENETICS INC.
PA
XX
XX Donoho G, Hilbun E, Turner CA, Friedrich G, Abuin A;
PI Zambrowicz B, Sands AT, Walke DW, Wilgowski NJ, Hu Y, Kieke JA;
PI Potter DG;
DR WPI; 2001-550185/61.
DR N-PSDB; AAS12623.
XX
XX Novel nucleic acid sequences encoding novel human proteins useful for
PT diagnosis, drug screening, clinical trial monitoring and treatment of
PT diseases and disorders.
PT
XX
XX Disclosure, Page 52-53; 60pp; English.
PS
XX
XX The invention relates to isolated nucleic acids encoding novel human
CC transferase proteins (NHP). The nucleic acids and proteins are useful for
CC diagnosis, drug screening, clinical trial monitoring and treatment of
CC diseases and disorders e.g. breast and prostate cancer. NHPs can also be
CC useful for augmenting the efficacy of chemotherapeutic agents used in
CC treatment of breast or prostate cancer. The nucleic acid is also useful
CC in NHP gene regulation, and as antisense primers in amplification
CC reactions of NHP gene sequences. NHPs are useful for producing
CC antibodies. The present sequence represents a novel human transferase
CC which has sequence similarity to N-acetyl-galactosaminyltransferase
XX
XX Sequence 366 AA;
SQ
Query Match 57.3%; Score 1877; DB 4; Length 366;
Best Local Similarity 100.0%; Pred. No. 1.5e-182;
Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MRRERKRLQAVALLVLAALVLLPNVGMALYRROPDGTGSGGAAYAPAGGSGSHRQK 60
DB 1 MRRERKRLQAVALLVLAALVLLPNVGMALYRROPDGTGSGGAAYAPAGGSGSHRQK 60

QY 61 KTFEFLDGGQLKXMDHKEAIRRDAQRVNGEGRPYPMTDABRVDOAYRENGENIYVSDK 120
 DB 61 KTFEFLDGGQLKXMDHKEAIRRDAQRVNGEGRPYPMTDABRVDOAYRENGENIYVSDK 120
 QY 121 ISANRLPDLRRHNCNCKRLETLPTNTSIIIPHNBSMSLLRTYHSVLRNSPPELVAEI 180
 DB 121 ISANRLPDLRRHNCNCKRLETLPTNTSIIIPHNBSMSLLRTYHSVLRNSPPELVAEI 180
 QY 181 VLVDSDREHLKKPLEDYVALFPYVRIATKKREGILRTKMLGASVATGDTVTFDLSHC 240
 DB 181 VLVDSDREHLKKPLEDYVALFPYVRIATKKREGILRTKMLGASVATGDTVTFDLSHC 240
 QY 241 EANNVNLPLLDRIARRNKTIVCPMIDVIDHDDFRYETQAGAMRGAFDWEMYKAIPIR 300
 DB 241 EANNVNLPLLDRIARRNKTIVCPMIDVIDHDDFRYETQAGAMRGAFDWEMYKAIPIR 300
 QY 301 PELQKADPSPDPSPVMAAGLPAVDRKFWELGQYDGLGELNGEGYEISFKV 353
 DB 301 PELQKADPSPDPSPVMAAGLPAVDRKFWELGQYDGLGELNGEGYEISFKV 353

RESULT 10

AA62600
 ID AA62600 standard; protein; 407 AA.

AC AA62600;
 XX
 DT 06-SEP-2001 (first entry)

DE Human acetyl galactosyl transferase 45.

XX Human; acetyl galactosyl transferase 45; GalNAc-T45; cancer; haemopathy;
 KM HIV infection; immunological disease; inflammation; gene therapy.

XX Homo sapiens.

PN WO200138545-A1.

PD 31-MAY-2001.

PF 20-NOV-2000; 2000MO-CN000473.

PR 24-NOV-1999; 99CN-00124100.

PA (BICR-) BICROAD GENE DEV LTD SHANGHAI.

PI Mao Y, Xie Y;

XX WPI, 2001-355942/37.

DR N-PSDB; AAA45701.

XX Human acetyl galactosyl transferase 45 and encoded polynucleotide, used
 PT in diagnosis and treatment of malignant tumors, hemopathy, human
 PT immunodeficiency virus infection, immunological diseases and
 PT inflammation.

XX

PS Claim 1; Page 20-21; 33pp; Chinese.

XX The present invention provides the protein and coding sequences of the
 CC human acetyl galactosyl transferase 45 (GalNAc-T45). The sequences can be
 CC used in the treatment of cancer, haemopathy, HIV infection, immunological
 CC diseases and inflammation. The present sequence is the protein of the
 CC invention
 CC

XX Sequence 407 AA.

Query Match 55.5%; Score 1818; DB 4; Length 407;

Best Local Similarity 78.5%; Pred. No. 2e-176;

Matches 317; Conservative 36; Mismatches 51; Indels 0; Gaps 0;

QY 200 MALPSPRIIRTKKREGLIRTRMLGASVATGDTVTFDLSCEANVMPLPLDLRIARRNK 259
 DB 1 MARFSKRIIVRTKREGLIRTRMLGASVATGDTVTFDLSCEANVMPLPLDLRIARRNK 60

QY 260 TIVCPMIDVIDHDDFRYETQAGDAMRGAFDWEMYKAIPIPELOKADSPDPSPVMA 319
 DB 61 TIVCPMIDVIDHDDFRYETQAGDAMRGAFDWEMYKAIPIPELOKADSPDPSPVMA 319
 QY 320 GLFAVDRKFWELGQYDGLGELNGEGYEISFKVMMCGRMEDIPCSRGHITRYKVPYK 379
 DB 121 GLFAVDRKFWELGQYDGLGELNGEGYEISFKVMMCGRMEDIPCSRGHITRYKVPYK 379
 QY 380 VPAGVSLAANLKRVAEVMDEVAEYIYORPEYRHLASDVAVOKLRSLNCKSPKWM 439
 DB 181 VPSGSLAANLKRVAEVMDEVAEYIYORPEYRHLASDVAVOKLRSLNCKSPKWM 439
 QY 440 TKIAMDPRKYPPEVPPAAWGEIRNVGTGLCADTGHGALGSPRLLEGCVRGGEAAMN 499
 DB 241 AAVAMDVPKYPPEVPPAAWGEIRNVANLGVDSHGATGTELRDLCVKDGSERTW 300
 QY 500 MQVFTTMRREDIRPGDPQHTKKCFPAIHTSPYLYDHSKNGNLMKTRDKTYHVP 559
 DB 301 EQLFTGMRREDIRPGDPQHTKKCFPAIHTSPYLYDHSKNGNLMKTRDKTYHVP 559
 QY 560 SGSCMDCSDRHIFMNTCPNSLTQOQLFEHTNSTVLEKFN 603
 DB 361 SNSCMDNPAKKIFMARCDPLSETOQWLFHINMTVLEKFN 404

RESULT 11

ABG08187
 ID ABG08187 standard; protein; 319 AA.

AC ABG08187;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #8178.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US008631.

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

PA (HYSE-) HYSEQ INC.

PI Dmanac RT, Liu C, Tang YT;

XX WPI, 2001-639362/73.

DR N-PSDB; AAS72374.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostic, forensic, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.

PS Claim 20; SEQ ID NO 38546; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying normal
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting and/or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food

CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 319 AA;

Query Match 42.8%; Score 1403.5; DB 4; Length 319;
 Best Local Similarity 90.8%; Pred. No. 3.5e-134;
 Matches 258; Conservative 4; Mismatches 19; Indels 3; Gaps 2;

DY 322 FAVDRKMFWEIGYDPLGEIWTGSGQYEISFVKWVGGRMEDIPCSRGHYRKVPYKVP 381
 12 FAEAKKCCMP-GGDELEAETSSSLFLAQWVMGGRMEDIPCSRGHYRKVPYKVP 70
 QY 382 AGVSLAR--NLKRAVAVMDEYAEYIYORRPEYHLSAGDVAVOKKLRSSLNCKSPFKMFM 439
 71 AGVSLARSKLKRVAEVMDEYAEYIYORRPEYHLSAGDVAVOKKLRSSLNCKSPFKMFM 130
 DY 440 TKIAMDLPKYPPVPPPAAMGEIRNVGTGICADTKRGALGSPRLRGCVRGREAAMNN 499
 131 TKIAMDLPKYPPVPPPAAMGEIRNVGTGICADTKRGALGSPRLRGCVRGREAAMNN 190
 QY 500 MOVFTFWRBDRIRGDPQHTKKCFDAISHTSPVTLVDCSMKNQMLKTKRDKTLVHPV 559
 191 MOVFTFWRBDRIRGDPQHTKKCFDAISHTSPVTLVDCSMKNQMLKTKRDKTLVHPV 250
 DY 560 SGSCMDCSESDHRIFNNTCNPSSLTQOVLFEHTNSTVLEKNNRN 603
 251 SGSCMDCSESDHRIFNNTCNPSSLTQOVLFEHTNSTVLEKNNRN 294

RESULT 12
 ADE25865
 ID ADE25865 standard; protein; 275 AA.

AC ADE25865;

DT 29-JAN-2004 (first entry)

DE Human N-acetylglucosaminyl transferase T10.

KM cytostatic; respiratory; antiasthmatic; antiinflammatory; endocrine; CNS;
 KM immunosuppressive; antiarthritic; antibacterial; vulnery;
 KM GalNAc-transferase lectin domain inhibitor;
 KM N-acetylglucosaminyl transferase; GalNAc-transferase; GalNAc-T4;
 KM GalNAc-T7; GalNAc-T3; GalNAc-transferase; tumour; cancer;
 KM lung disease; mucous accumulation; asthma; chronic bronchitis;
 KM smoker's lung; cystic fibrosis; exocrine gland disease;
 KM mucin secretion decrease; Sjogren's syndrome; dry mouth;
 KM selection-mediated leukocyte trafficking; autoimmunity; arthritis;
 KM leukaemia; lymphoma; immunosuppression; sepsis; wound healing;
 KM inflammation; mucin secretion inhibition; hypersecretion;
 KM mucin accumulation; chronic obstructive respiratory pulmonary disease;
 KM asthma; lectin domain inhibitor; GalNAc-glycopeptide;
 KM Galbeta1-3GalNAc-glycopeptide; UDP-Gal donor; glycosyltransferase;
 KM core 1 beta1,3-galactosyltransferase; alpha2,6-sialyltransferase;
 KM N-acetylglucosaminyl transferase 4AP24 glycopeptide; O-glycosylation;
 KM breast cancer; human; GalNAcT10.

OS Homo sapiens.

PN US2003186850-A1.

PD 02-OCT-2003.

XX 12-NOV-2002; 2002US-00292896.

XX 11-MAY-2000; 2000US-0203331P.

PR 10-MAY-2001; 2001WO-DX000328.

PR 08-NOV-2002; 2002US-0425204P.

PA (GLYC-) GLYCOSYLV APS.

XX Clausen H, Bennett EP, Haesean H, Reis CA;

XX WPI; 2003-831259/77.

PT Modulating N-acetylglucosaminyltransferase functions comprises

PS administration of an N-acetylglucosaminyltransferase lectin domain

XX inhibitor.

XX Disclosure; Page 11; 65pp; English.

CC The invention describes a method of modulating polypeptide N-
 CC acetylglucosaminyl (GalNAc)-transferase functions comprising
 CC administration of a GalNAc-transferase lectin domain inhibitor (I). (I)
 CC is used for modulating functions of GalNAc-transferase (e.g. GalNAc-T4,
 CC GalNAc-T7, GalNAc-T2, and GalNAc-T3); for inhibiting at least one lectin
 CC domain of the GalNAc-transferase and modulating the function mediated by
 CC the domain for preparing a medicament for treating tumours and cancers;
 CC lung diseases associated with mucous accumulation (e.g. asthma, chronic
 CC bronchitis, smoker's lung, and cystic fibrosis), diseases of exocrine
 CC glands associated with increased or decreased mucin secretion (e.g.
 CC Sjogren's syndrome and dry mouth), disorders associated with
 CC dysregulation of selectin-mediated leukocyte trafficking (e.g.
 CC autoimmunity, arthritis, leukaemia, lymphoma, immunosuppression,
 CC sepsis, wound healing, acute and chronic inflammation); inhibiting mucin
 CC secretion; and inhibiting hypersecretion and accumulation of mucin in the
 CC lungs of mammals (e.g. humans) suffering from chronic obstructive
 CC respiratory pulmonary diseases, asthma, and cystic fibrosis. (I) is
 CC effective in inhibiting at least one lectin domain of the GalNAc-
 CC transferase and modulating the inhibited function mediated by the domain
 CC (preferably GalNAc-glycopeptide, Galbeta1-3GalNAc-glycopeptide, or
 CC peptide specificity of the transferase, or UDP-Gal donor substrate
 CC specificity). (I) selectively inhibits at least one member of the GalNAc-
 CC transferase family without inhibiting other glycosyltransferases
 CC (preferably core 1 beta1,3-galactosyltransferase, alpha2,6-
 CC sialyltransferase, or glycosyltransferase functioning in the O-
 CC glycosylation pathway). This is the amino acid sequence of human N-
 CC acetylglucosaminyl transferase T10.

SQ Sequence 275 AA;

Query Match 42.6%; Score 1396; DB 7; Length 275;
 Best Local Similarity 93.7%; Pred. No. 1.6e-133;
 Matches 253; Conservative 4; Mismatches 11; Indels 2; Gaps 1;

DY 336 DPGLEIWTGSGQYEISFVKWVGGRMEDIPCSRGHYRKVPYKVPAGVSLA--RLMKRV 393
 6 DGELEETSSSLFLAQWVMGGRMEDIPCSRGHYRKVPYKVPAGVSLAVRTIKRV 65
 DB 394 AAVWMDYAEYIYORRPEYHLSAGDVAVOKKLRSSLNCKSPFKMFMTKIAMDLPKYPPV 453
 66 AAVWMDYAEYIYORRPEYHLSAGDVAVOKKLRSSLNCKSPFKMFMTKIAMDLPKYPPV 125
 DY 454 EPPPAAMGEIRNVGTGICADTKRGALGSPRLRGCVRGREAAMNNQVFTFWRBDRIR 513
 126 EPPPAAMGEIRNVGTGICADTKRGALGSPRLRGCVRGREAAMNNQVFTFWRBDRIR 185
 DB 514 GDPQHTKKCFDAISHTSPVTLVDCSMKNQMLKTKRDKTLVHPVSGSCMDCSESDHRI 573
 186 GDPQHTKKCFDAISHTSPVTLVDCSMKNQMLKTKRDKTLVHPVSGSCMDCSESDHRI 245
 DY 574 FNNTCNPSSLTQOVLFEHTNSTVLEKNNRN 603
 246 FNNTCNPSSLTQOVLFEHTNSTVLEKNNRN 275

RESULT 13
ADQ17613 standard; protein: 276 AA.
ID ADQ17613
XX
AC ADQ17613;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated protein - SEQ ID 430.
XX
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.
XX
OS Homo sapiens.
XX
PN WO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PI (PROT-) PROTEIN DESIGN LABS INC.
XX
PA Aziz N, Ginsburg WM, Zlotnik A;
XX
DR WPI; 2004-441208/41.
XX
PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
PS Example 2; SEQ ID NO 430; 210bp; English.
XX
CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC protein of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 276 AA;
XX
Query Match 42.5%; Score 1392; DB 8; Length 276;
Best Local Similarity 97.3%; Pred. No. 4.2e-133;
Matches 250; Conservative 3; Mismatches 2; Indels 2; Gaps 1;
XX
OY 349 ISFVMMCGGRMEDIPCSRVGHIYRKVPYKVPAGVSLA--RNLKRYAEVMMDEYAYIY 406
:::|||||
DB 20 LAMQVMMCGGRMEDIPCSRVGHIYRKVPYKVPAGVSLA|KRYAEVMMDEYAYIY 79
XX
OY 407 QRRREYHLSAGDVAVOKLRSLSLCKSPKFMFTKIAMDLPKFYPVPEPPAAAMGEIRNV 466
80 QRRREYHLSAGDVAVOKLRSLSLCKSPKFMFTKIAMDLPKFYPVPEPPAAAMGEIRNV 139
XX
OY 467 GTGICADTKXGALGSPRLRGGCVARGGEAAMNNQVTFTRREDIRGDDQHTKKCFDA 526
140 GTGICADTKXGALGSPRLRGGCVARGGEAAMNNQVTFTRREDIRGDDQHTKKCFDA 199
XX
OY 527 ISHTSPVTLVDCHSMKGNOLMKYRKDXTLYHPVSGSCMDSESDDHRIFMKTCNPSSLTQQ 586
200 ISHTSPVTLVDCHSMKGNOLMKYRKDXTLYHPVSGSCMDSESDDHRIFMKTCNPSSLTQQ 259
DB 587 WLFHTNSTVLEKFNRN 603
OY

DB 260 WLFHTNSTVLEKFNRN 276
|||||
RESULT 14
ABG08189 standard; protein: 335 AA.
ID ABG08189
XX
AC ABG08189;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #8180.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
XX
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX
DR N-PSDB; AAS72376.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensic, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 38548; 103bp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probe, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for creating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensic, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 335 AA;
XX
Query Match 42.3%; Score 1385.5; DB 4; Length 335;
Best Local Similarity 87.5%; Pred. No. 2.7e-132;
Matches 253; Conservative 0; Mismatches 1; Indels 35; Gaps 1;
XX
OY 350 SFRVMMCGGRMEDIPCSRVGHIYRKVPYKVPAGVSLA|NLKRYAEVMMDEYAYIYORR 409
47 SFRVMMCGGRMEDIPCSRVGHIYRKVPYKVPAGVSLA|NLKRYAEVMMDEYAYIYORR 106
DB

Search completed: December 20, 2004, 14:09:34
Job time : 164 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2004, 15:26:33 ; Search time 40 Seconds

(Without alignments)
999.744 Million cell updates/sec

Title: US-10-001-851-2

Perfect score: 3278
Sequence: 1 MRKKERKLQAVLVLAIV.....TQQWLFPHNTSTYLEKRN 603

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 100 summaries

Database :

1: Issued Patents AA:*
2: /cgn2_6/ptodata/1/iaa/5A.COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PTCONS.COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3278	100.0	603	4	US-09-795-926-43
2	3242.5	98.9	631	4	US-09-795-926-41
3	2771	84.5	506	4	US-09-795-926-31
4	2746.5	83.8	535	4	US-09-795-926-39
5	1877	57.3	366	4	US-09-795-926-39
6	1370	41.8	269	4	US-09-795-926-27
7	1311	40.0	321	4	US-09-795-926-35
8	1115	34.0	517	4	US-08-967-508-19
9	1115	34.0	517	3	US-08-967-508-19
10	1115	34.0	517	5	PCT-US94-02552-19
11	1115	34.0	559	2	US-08-967-508-9
12	1115	34.0	559	3	US-08-967-506-9
13	1115	34.0	559	5	PCT-US94-02552-9
14	1026	31.3	578	4	US-09-217-306B-2
15	1024	31.2	560	4	US-09-217-306B-10
16	990	30.2	240	4	US-09-795-926-33
17	988	30.1	209	4	US-09-795-926-37
18	971	29.6	552	4	US-10-140-002-196
19	941	28.7	302	4	US-09-270-767-33733
20	926	28.2	639	3	US-09-376-856-2
21	920	28.1	638	3	US-09-347-488-2
22	916	27.9	633	2	US-08-648-298-2
23	804	24.5	224	4	US-09-795-926-23
24	744	22.7	603	4	US-09-973-457-2
25	675	20.6	289	4	US-09-270-767-45334
26	561	17.1	238	4	US-09-270-767-46411
27	503.5	15.4	366	4	US-09-270-767-43543

28	483	14.7	143	4	US-09-795-926-21	Sequence 21, Appl
29	481	14.7	112	4	US-09-795-926-25	Sequence 25, Appl
30	450.5	13.7	149	4	US-09-973-457-6	Sequence 6, Appl
31	428.5	13.1	161	4	US-09-270-767-58910	Sequence 58910, A
32	287	8.8	122	4	US-09-513-996C-7570	Sequence 7570, Ap
33	277.5	8.5	148	4	US-09-270-767-32872	Sequence 32872, A
34	241.5	7.4	313	4	US-09-270-767-46209	Sequence 46209, A
35	234.5	7.2	300	4	US-09-270-767-43809	Sequence 43809, A
36	228.5	7.0	220	4	US-09-270-767-61773	Sequence 61773, A
37	218.5	6.7	72	4	US-09-270-767-61989	Sequence 61989, A
38	168	5.1	73	4	US-09-270-767-33350	Sequence 33350, A
39	168	5.1	73	4	US-09-270-767-33341	Sequence 33341, A
40	167	5.1	101	4	US-09-270-767-33441	Sequence 33441, A
41	167	5.1	101	4	US-09-270-767-48658	Sequence 48658, A
42	133	4.1	424	4	US-09-134-000C-5967	Sequence 5967, Ap
43	129	3.9	88	4	US-09-270-767-33601	Sequence 33601, A
44	129	3.9	88	4	US-09-270-767-48818	Sequence 48818, A
45	128	3.9	135	4	US-09-973-457-5	Sequence 5, Appl
46	127	3.9	436	4	US-09-328-352-4416	Sequence 4416, Ap
47	125.5	3.8	965	4	US-09-437-277-3	Sequence 3, Appl
48	124	3.8	702	4	US-09-437-277-1	Sequence 1, Appl
49	123.5	3.8	534	4	US-09-134-000C-5087	Sequence 5087, Ap
50	123	3.8	721	4	US-09-107-532A-6889	Sequence 6889, Ap
51	121.5	3.7	1056	4	US-09-134-000C-5086	Sequence 5086, Ap
52	119.5	3.6	492	2	US-08-468-812-4	Sequence 4, Appl
53	119.5	3.6	492	2	US-08-468-812-7	Sequence 7, Appl
54	119.5	3.6	492	3	US-08-590-563-4	Sequence 4, Appl
55	119.5	3.6	492	3	US-08-590-563-7	Sequence 7, Appl
56	119.5	3.6	492	4	US-09-770-621-4	Sequence 4, Appl
57	119.5	3.6	492	4	US-09-770-621-7	Sequence 7, Appl
58	119.5	3.6	492	4	US-09-235-832-7	Sequence 4, Appl
59	119.5	3.6	492	4	US-09-235-832-4	Sequence 4, Appl
60	118	3.6	480	2	US-08-468-812-5	Sequence 5, Appl
61	118	3.6	480	3	US-08-590-563-5	Sequence 5, Appl
62	118	3.6	480	4	US-09-770-621-5	Sequence 5, Appl
63	118	3.6	480	4	US-09-235-832-5	Sequence 5, Appl
64	116.5	3.6	88	4	US-09-270-767-33890	Sequence 33890, A
65	116.5	3.6	88	4	US-09-270-767-49107	Sequence 49107, A
66	112.5	3.4	909	3	US-09-425-383-2	Sequence 2, Appl
67	111.5	3.4	491	2	US-08-468-812-8	Sequence 8, Appl
68	111.5	3.4	491	3	US-08-590-563-8	Sequence 8, Appl
69	111.5	3.4	491	4	US-09-770-621-8	Sequence 8, Appl
70	111.5	3.4	491	4	US-09-770-621-8	Sequence 8, Appl
71	107.5	3.3	507	4	US-09-235-832-8	Sequence 8, Appl
72	107	3.3	270	4	US-09-130-337A-25	Sequence 25, Appl
73	107	3.3	270	4	US-09-495-406-25	Sequence 25, Appl
74	107	3.3	270	4	US-09-816-028A-39	Sequence 39, Appl
75	105.5	3.2	187	4	US-10-303-162-32	Sequence 32, Appl
76	105.5	3.2	136	4	US-09-543-681A-6535	Sequence 6535, Ap
77	105	3.2	127	1	US-08-392-828C-39	Sequence 39, Appl
78	105	3.2	127	3	US-09-330-945-39	Sequence 39, Appl
79	105	3.2	324	1	US-08-597-236-10	Sequence 10, Appl
80	105	3.2	324	1	US-08-746-682A-10	Sequence 10, Appl
81	102.5	3.1	540	1	US-08-378-761A-77	Sequence 77, Appl
82	102.5	3.1	540	1	US-08-485-886-77	Sequence 77, Appl
83	102	3.1	317	4	US-09-583-110-4849	Sequence 4849, Ap
84	101.5	3.1	395	4	US-08-635-552A-4	Sequence 4, Appl
85	101.5	3.1	419	2	US-08-270-81-2	Sequence 2, Appl
86	101.5	3.1	419	4	US-09-146-693-2	Sequence 2, Appl
87	101.5	3.1	419	4	US-08-675-499A-5	Sequence 5, Appl
88	101.5	3.1	419	4	US-08-812-008-5	Sequence 5, Appl
89	101	3.1	328	4	US-09-583-110-3554	Sequence 3554, Ap
90	100	3.0	331	4	US-09-634-238-247	Sequence 247, Ap
91	99.5	3.0	303	4	US-09-495-406-17	Sequence 17, Appl
92	99.5	3.0	303	4	US-09-816-028A-29	Sequence 29, Appl
93	99.5	3.0	303	4	US-10-303-162-22	Sequence 22, Appl
94	99	3.0	341	4	US-09-328-352-7178	Sequence 7178, Ap
95	98	3.0	301	4	US-09-495-406-15	Sequence 15, Appl
96	98	3.0	301	4	US-09-816-028A-27	Sequence 27, Appl
97	98	3.0	301	4	US-10-303-162-27	Sequence 27, Appl
98	96.5	2.9	623	1	US-08-653-740-7	Sequence 7, Appl
99	96.5	2.9	623	2	US-09-073-594-7	Sequence 7, Appl
100	96.5	2.9	623	3	US-09-275-925-7	Sequence 7, Appl

ALIGNMENTS

```
RESULT 1
US-09-795-926-43
; Sequence 43, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedlich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sander, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Kleke, James Alvin
; APPLICANT: Poter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; NUMBER OF SEQ. ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ. ID NO 43
; LENGTH: 603
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-795-926-43

Query Match      100.0%; Score 3278; DB 4; Length 603;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 603; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MRREKRLQAVLVLAALVLPVGLMALYREROPGTGGSGAAYAPAGGSGSHRQK 60
DB      1 MRREKRLQAVLVLAALVLPVGLMALYREROPGTGGSGAAYAPAGGSGSHRQK 60
QY      61 KTFELGGQGLKQMDHKEAIRDAQVNGEQGRPYMTDAERVDQAYRENGFNIVYSDK 120
DB      61 KTFELGGQGLKQMDHKEAIRDAQVNGEQGRPYMTDAERVDQAYRENGFNIVYSDK 120
QY      121 ISLNRSIPDTRHPCNSKRYLETLPNTSIIIPPHNEGWSLLRTVSHVLRSPPELVAET 180
DB      121 ISLNRSIPDTRHPCNSKRYLETLPNTSIIIPPHNEGWSLLRTVSHVLRSPPELVAET 180
QY      181 VLVDDFSDREHLKKPLEDYVALPFSVRLRTKKREGILTRMLGASVATGDTVTFDLSHC 240
DB      181 VLVDDFSDREHLKKPLEDYVALPFSVRLRTKKREGILTRMLGASVATGDTVTFDLSHC 240
QY      241 EAVVNWLPPLLDRIARNRKTIVCEMIDVIDHDDFRYETQAGDAMRGAFDWMYYKRLPIP 300
DB      241 EAVVNWLPPLLDRIARNRKTIVCEMIDVIDHDDFRYETQAGDAMRGAFDWMYYKRLPIP 300
QY      301 PELQKADSPDPESPVAAGGLFAVDRKFWELGSDGLIWGGEQYEISFKYMMCGGRM 360
DB      301 PELQKADSPDPESPVAAGGLFAVDRKFWELGSDGLIWGGEQYEISFKYMMCGGRM 360
QY      361 EDIPCSRVGHIYRYKYVPYKYPAGVSLARNLKRVAVEMWDEYAEITYORREYRHLSDGV 420
DB      361 EDIPCSRVGHIYRYKYVPYKYPAGVSLARNLKRVAVEMWDEYAEITYORREYRHLSDGV 420
```

```
QY      421 AVQKLRSSLNCKSPFWMTKIAMDLPKFYPPVEPPAAAMGEIRNVGTGICADTKGALG 480
DB      421 AVQKLRSSLNCKSPFKPMFKIAMDLPKFYPPVEPPAAAMGEIRNVGTGICADTKGALG 480
QY      481 SPLRLGCVRGREGAAMNNQVFTFWREDIRGDPQHTKKCFDAISHTSPVTLVDCHS 540
DB      481 SPLRLGCVRGREGAAMNNQVFTFWREDIRGDPQHTKKCFDAISHTSPVTLVDCHS 540
QY      541 MKGNQLMKYRKDITLYHPVSGSCMDCSDHRIFNATCNPSLTOQMLEHTNSTVLEKF 600
DB      541 MKGNQLMKYRKDITLYHPVSGSCMDCSDHRIFNATCNPSLTOQMLEHTNSTVLEKF 600
QY      601 NRN 603
DB      601 NRN 603

RESULT 2
US-09-795-926-41
; Sequence 41, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedlich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sander, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Kleke, James Alvin
; APPLICANT: Poter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; NUMBER OF SEQ. ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ. ID NO 41
; LENGTH: 631
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-795-926-41

Query Match      98.9%; Score 3242.5; DB 4; Length 631;
Best Local Similarity 95.4%; Pred. No. 0;
Matches 601; Conservative 0; Mismatches 0; Indels 29; Gaps 1;

QY      1 MRREKRLQAVLVLAALVLPVGLMALYREROPGTGGSGAAYAPAGGSGSHRQK 60
DB      1 MRREKRLQAVLVLAALVLPVGLMALYREROPGTGGSGAAYAPAGGSGSHRQK 60
QY      61 KTFELGGQGLKQMDHKEAIRDAQVNGEQGRPYMTDAERVDQAYRENGFNIVYSDK 120
DB      61 KTFELGGQGLKQMDHKEAIRDAQVNGEQGRPYMTDAERVDQAYRENGFNIVYSDK 120
QY      121 ISLNRSIPDTRHPCNSKRYLETLPNTSIIIPPHNEGWSLLRTVSHVLRSPPELVAET 180
DB      121 ISLNRSIPDTRHPCNSKRYLETLPNTSIIIPPHNEGWSLLRTVSHVLRSPPELVAET 180
QY      181 VLVDDFSDREHLKKPLEDYVALPFSVRLRTKKREGILTRMLGASVATGDTVTFDLSHC 240
DB      181 VLVDDFSDREHLKKPLEDYVALPFSVRLRTKKREGILTRMLGASVATGDTVTFDLSHC 240
```

QY 241 EANNVNLPLLDRIANRRTIVCPMIDVIDHDDFRVETQAGDAMRGAFDWEYTKRPIIP 300
DB 241 EANNVNLPLLDRIANRRTIVCPMIDVIDHDDFRVETQAGDAMRGAFDWEYTKRPIIP 300
QY 301 PELQKADPSDPFESSPVWAGGLFVADRKMFWEELGSGYDPLGELIWGGEYEISFK----- 352
DB 301 PELQKADPSDPFESSPVWAGGLFVADRKMFWEELGSGYDPLGELIWGGEYEISFKGLHMLPRL 360
QY 353 -----VMMCGRMEDIPCRRGHYRKVYKVPKVPAGVSLANLK 391
DB 361 VNSNMQAVFLPPAPNNMALQVMMCGRMEDIPCRRGHYRKVYKVPKVPAGVSLANLK 420
QY 392 RVAEVMWDEYAEYIYORREYRHLASGVAVQKKLRSSLNCKSFKFMFTKIAMDLPKFPY 451
DB 421 RVAEVMWDEYAEYIYORREYRHLASGVAVQKKLRSSLNCKSFKFMFTKIAMDLPKFPY 480
QY 452 PVEPPAAWGEIRNVGTGLCADTKHGALGSPRLREGCVGRGGAAMNNQVFTFWREDI 511
DB 481 PVEPPAAWGEIRNVGTGLCADTKHGALGSPRLREGCVGRGGAAMNNQVFTFWREDI 540
QY 512 RRGDPQHTTKFCFCDALSHTSPTLYDCHSMKGNQJMKYRKDKTLVHPVSGSCMDCSESDH 571
DB 541 RRGDPQHTTKFCFCDALSHTSPTLYDCHSMKGNQJMKYRKDKTLVHPVSGSCMDCSESDH 600
QY 572 RIFMNTCNPSLTOQMLFEHTNSTVLEKFN 601
DB 601 RIFMNTCNPSLTOQMLFEHTNSTVLEKFN 630

RESULT 3
US-09-795-926-31
; Sequence 31, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuhin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 506
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-795-926-31

Query Match 84.5%; Score 2771; DB 4; Length 506;
Best Local Similarity 100.0%; Pred. No. 2.7e-289; Indels 0; Gaps 0;
Matches 506; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 158 WSLIARTVHSVNLNRSBPPELVAEIVLVDDPSDEHLLKPLLEDYVALPSPVRIILTKRREGI 217
DB 61 WSLIARTVHSVNLNRSBPPELVAEIVLVDDPSDEHLLKPLLEDYVALPSPVRIILTKRREGI 120
QY 218 ITRMIGASVATGDTITPLDISHCEANVNLPLLDRIANRRTIVCPMIDVIDHDDFRYE 277
DB 121 ITRMIGASVATGDTITPLDISHCEANVNLPLLDRIANRRTIVCPMIDVIDHDDFRYE 180
QY 278 TQAGDMRGAFFMWEYTKRPIIPPELOKADPSDPFESSPVWAGGLFVADRKMFWEELGSGYD 337
DB 181 TQAGDMRGAFFMWEYTKRPIIPPELOKADPSDPFESSPVWAGGLFVADRKMFWEELGSGYD 240
QY 338 GLEIWMGEYEISFKVMCGRMEDIPCRRVGHYRKVYKVPKVPAGVSLANLKRAVAEYV 397
DB 241 GLEIWMGEYEISFKVMCGRMEDIPCRRVGHYRKVYKVPKVPAGVSLANLKRAVAEYV 300
QY 398 MDEYAEYIYORREYRHLASGVAVQKKLRSSLNCKSFKFMFTKIAMDLPKFPYVPEPPA 457
DB 301 MDEYAEYIYORREYRHLASGVAVQKKLRSSLNCKSFKFMFTKIAMDLPKFPYVPEPPA 360
QY 458 AAMGEIRNVGTGLCADTKHGALGSPRLREGCVGRGGAAMNNQVFTFWREDIRGDDQ 517
DB 361 AAMGEIRNVGTGLCADTKHGALGSPRLREGCVGRGGAAMNNQVFTFWREDIRGDDQ 420
QY 518 HTKFCFCDALSHTSPTLYDCHSMKGNQJMKYRKDKTLVHPVSGSCMDCSESDHIFMNT 577
DB 421 HTKFCFCDALSHTSPTLYDCHSMKGNQJMKYRKDKTLVHPVSGSCMDCSESDHIFMNT 480
QY 578 CNPSSLTOQMLFEHTNSTVLEKFN 603
DB 481 CNPSSLTOQMLFEHTNSTVLEKFN 506

RESULT 4
US-09-795-926-29
; Sequence 29, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuhin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 535
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-795-926-29

Query Match 83.8%; Score 2746.5; DB 4; Length 535;
Best Local Similarity 94.6%; Pred. No. 1.3e-286;
Matches 506; Conservative 0; Mismatches 0; Indels 29; Gaps 1;

```

QY 98 MTDARVDAQYRENGFNIVYSDKISLNRSLPDIRHPNCNKRYLETLPNTSIIIPHNHG 157
DB 1 MTDARVDAQYRENGFNIVYSDKISLNRSLPDIRHPNCNKRYLETLPNTSIIIPHNHG 60
QY 158 WSSLLRTVSHVLRNRPPELVAEIVLVDDFSDREHLKKPLEDMALPFSVAILTKKREGI 217
DB 61 WSSLLRTVSHVLRNRPPELVAEIVLVDDFSDREHLKKPLEDMALPFSVAILTKKREGI 120
QY 218 IRRFMIGASVATGDTVITFLDSHCEANVMPLPILLDIRIARRKTIIVCPMIDVIVHDDPRYE 277
DB 121 IRRFMIGASVATGDTVITFLDSHCEANVMPLPILLDIRIARRKTIIVCPMIDVIVHDDPRYE 180
QY 278 TQAGDAMRGAFDWMYKRIPIPELQKADPSDFPESPVNAGGLFVDRKFMWELGSDYD 337
DB 181 TQAGDAMRGAFDWMYKRIPIPELQKADPSDFPESPVNAGGLFVDRKFMWELGSDYD 240
QY 338 GLEINGGEQYEISFK-----VWCGGRMEDIPCSRY 368
DB 241 GLEINGGEQYEISFKHMLPRLVSNMPOAVFLPRAPNMALQVWCGGRMEDIPCSRY 300
QY 369 GHYRKVPYKVPAGVSLAHLKRVAVVMDEYAEYIYORRPEYRHLSDGVAVOKKLRS 428
DB 301 GHYRKVPYKVPAGVSLAHLKRVAVVMDEYAEYIYORRPEYRHLSDGVAVOKKLRS 360
QY 429 SLNCKSFKWMFKIANDLPKFPYVBPAPAAWGEIRNVGTGLCADTKGALGSPRLREGC 488
DB 361 SLNCKSFKWMFKIANDLPKFPYVBPAPAAWGEIRNVGTGLCADTKGALGSPRLREGC 420
QY 489 VRGRGEAAMNMVOFTTWMEDIRPGDPQHTKKFCFPAISHTSPVTLYDCHSKGNQJMK 548
DB 421 VRGRGEAAMNMVOFTTWMEDIRPGDPQHTKKFCFPAISHTSPVTLYDCHSKGNQJMK 480
QY 549 YRDKKTLVHPVSGSCMDCESEDRIFMNTCNPSLSLTOQMLEHTNSTVLEKFNRN 603
DB 481 YRDKKTLVHPVSGSCMDCESEDRIFMNTCNPSLSLTOQMLEHTNSTVLEKFNRN 535

RESULT 5
US-09-795-926-39
; Sequence 39, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedlich, Glenn
; APPLICANT: Abulin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Poter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 366
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-795-926-39

```

```

Query Match 57.3%; Score 1877; DB 4; Length 366;
Best Local Similarity 100.0%; Pred. No. 3.5e-193;
Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKKEKRLIQAVLVTAALVLLPNTGLMALYEROPDGTGGSGAAVAPAAAGGSHSRK 60
DB 1 MRKKEKRLIQAVLVTAALVLLPNTGLMALYEROPDGTGGSGAAVAPAAAGGSHSRK 60
QY 61 KTFPLDGGOKLQDMHKEAIRDAQVNGEGGRPYPMIDARVDAQYRENGFNIVYSDK 120
DB 61 KTFPLDGGOKLQDMHKEAIRDAQVNGEGGRPYPMIDARVDAQYRENGFNIVYSDK 120
QY 121 ISLNRSLPDIRHPNCNKRYLETLPNTSIIIPHNHGWSLLRTVSHVLRNRPPELVAEI 180
DB 121 ISLNRSLPDIRHPNCNKRYLETLPNTSIIIPHNHGWSLLRTVSHVLRNRPPELVAEI 180
QY 181 VLVDPSDREHLKKPLEDMALPFSVAILTKKREGIITRMLGASVATGDTVITFLDSHC 240
DB 181 VLVDPSDREHLKKPLEDMALPFSVAILTKKREGIITRMLGASVATGDTVITFLDSHC 240
QY 241 EAVVNMPLPILLDIRIARRKTIIVCPMIDVIVHDDPRYETQAGDAMRGAFDWMYKRIPI 300
DB 241 EAVVNMPLPILLDIRIARRKTIIVCPMIDVIVHDDPRYETQAGDAMRGAFDWMYKRIPI 300
QY 301 PELQKADPSDFPESPVNAGGLFVDRKFMWELGSDYDGLIENGGEQYEISFKY 353
DB 301 PELQKADPSDFPESPVNAGGLFVDRKFMWELGSDYDGLIENGGEQYEISFKY 353

RESULT 6
US-09-795-926-27
; Sequence 27, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedlich, Glenn
; APPLICANT: Abulin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Poter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 269
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-795-926-27

Query Match 41.8%; Score 1370; DB 4; Length 269;
Best Local Similarity 100.0%; Pred. No. 9.1e-139;
Matches 256; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 MTDARVDAQYRENGFNIVYSDKISLNRSLPDIRHPNCNKRYLETLPNTSIIIPHNHG 157
DB 1 MTDARVDAQYRENGFNIVYSDKISLNRSLPDIRHPNCNKRYLETLPNTSIIIPHNHG 60

```

QY 158 MSSLLRTVHSLNRPPELVAEIYVDFSDREHLKKPLEDYALPFSVAILRTKKEGL 217
 DB 61 MSSLLRTVHSLNRPPELVAEIYVDFSDREHLKKPLEDYALPFSVAILRTKKEGL 120
 QY 218 IRTRMGASVATGVDVITFLDSCGANVNLPLLDRIARNRKTIICPMIDIVIHDDFRYE 277
 DB 121 IRTRMGASVATGVDVITFLDSCGANVNLPLLDRIARNRKTIICPMIDIVIHDDFRYE 180
 QY 278 TQAGDMRGAFDWMYKRIPIPELOKADSPDFESPVAAGLFAVDRKMFELGSDP 337
 DB 181 TQAGDMRGAFDWMYKRIPIPELOKADSPDFESPVAAGLFAVDRKMFELGSDP 240
 QY 338 GLEIWGEQYXISFKV 353
 DB 241 GLEIWGEQYXISFKV 256

RESULT 7
 US-09-795-926-35
 / Sequence 35, Application US/09795926
 / Patent No. 655669
 / GENERAL INFORMATION:
 / APPLICANT: Donoho, Gregory
 / APPLICANT: Hilbun, Erin
 / APPLICANT: Turner, C. Alexander Jr.
 / APPLICANT: Friedrich, Glenn
 / APPLICANT: Abuin, Alejandro
 / APPLICANT: Zambrowicz, Brian
 / APPLICANT: Sande, Arthur T.
 / APPLICANT: Wilganowski, Nathaniel L.
 / APPLICANT: Hu, Yi
 / APPLICANT: Kieke, James Alvin
 / APPLICANT: Potter, David George
 / TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
 / FILE REFERENCE: LEX-0144-USA
 / CURRENT APPLICATION NUMBER: US/09/795,926
 / PRIOR APPLICATION NUMBER: US 60/185,920
 / PRIOR FILING DATE: 2000-02-29
 / PRIOR APPLICATION NUMBER: US 60/186,558
 / PRIOR FILING DATE: 2000-03-02
 / PRIOR APPLICATION NUMBER: US 60/191,849
 / PRIOR FILING DATE: 2000-03-24
 / NUMBER OF SEQ ID NOS: 47
 / SOFTWARE: FastSeq for windows Version 4.0
 / SEQ ID NO 35
 / LENGTH: 321
 / TYPE: PRT
 / ORGANISM: homo sapiens
 / US-09-795-926-35

Query Match 40.0%; Score 1311; DB 4; Length 321;
 Best Local Similarity 99.2%; Pred. No. 2,8e-112;
 Matches 251; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 241 EANVNLPLLDRI 253
 DB 241 EANVNLPLLDRI 253

RESULT 8
 US-08-967-508-19
 / Sequence 19, Application US/08967508
 / Patent No. 5910570
 / GENERAL INFORMATION:
 / APPLICANT: The Upjohn Company
 / TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GalNAc:
 / NUMBER OF SEQUENCES: 19
 / CORRESPONDENCE ADDRESS:
 / ADDRESSEE: Pharmacia & Upjohn Company, Intellectual
 / ADDRESSEE: Property Legal Services
 / STREET: 301 Henrietta Street
 / CITY: Kalamazoo
 / STATE: Michigan
 / COUNTRY: USA
 / ZIP: 49001
 / COMPUTER READABLE FORM:
 / MEDIUM TYPE: Floppy disk
 / COMPUTER: IBM PC compatible
 / OPERATING SYSTEM: PC-DOS/MS-DOS
 / SOFTWARE: Patentin Releasee #1.0, Version #1.25
 / CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: US/08/967,508
 / FILING DATE:
 / CLASSIFICATION: 435
 / PRIOR APPLICATION DATA:
 / APPLICATION NUMBER: 08/602,830
 / FILING DATE: 13 No. 5910570ember 1995
 / ATTORNEY/AGENT INFORMATION:
 / NAME: Darnley Jr., James D.
 / REGISTRATION NUMBER: 33,673
 / REFERENCE/DOCKET NUMBER: 4755.P CP
 / TELECOMMUNICATION INFORMATION:
 / TELEPHONE: 616-833-2210
 / TELEFAX: 616-833-8897
 / TELEX: 224401
 / INFORMATION FOR SEQ ID NO: 19:
 / SEQUENCE CHARACTERISTICS:
 / LENGTH: 517 amino acids
 / TYPE: amino acid
 / STRANDEDNESS: single
 / TOPOLOGY: linear
 / MOLECULE TYPE: protein
 / US-08-967-508-19

Query Match 34.0%; Score 1115; DB 2; Length 517;
 Best Local Similarity 43.6%; Pred. No. 8.3e-111;
 Matches 230; Conservative 89; Mismatches 170; Indels 38; Gaps 16;

```

Db 253 LFSIDRDFOBIETGADMGIMGEMLEISFRIMQCGTLEIYTCSHVGHVFRKATPYTF 312
QY 381 PAGVS--LARNLRVAEVMWDEYAEYIYQRRPEYRHLSAGDVAVQKLRSSLNCKSPKMF 438
Db 313 PGGTGQIINKNNRRLAEVWMDDEFKNFFYIISPGVTKYDGDISSRLGLRHLQCRPFWSY 372
QY 439 MTKIAMD--LPKFPYPPVEPPAAAGWGEIRNVGTGLCADTKHGLGSPRLLEGVGRGEAA 496
Db 373 LENIYDPSQIPRHY-----FSLGEIRNVETWQCLDNNAKENEKXGIFNC--HGWG--- 421
QY 497 WNNMQUFTFTWRREDIRGDPQHTKKFCFPAISHTSPVTLYDCHSMKGNOLMYKRXDK--TL 555
Db 422 --GNQVFSYANKKEIRTD-----LCIDVSKLNGVPTMLKCHHLKGNOLMEYDVPVXLTL 473
QY 556 YHPSGSCMD-CSESDHRI-FMNTCNPSLTOQMLFEHTNSTVLEKF 600
Db 474 QHNSNQCLDKATDEDSQVPSIRDCS-GSRSQOMLNR--NVTLPEIF 517

```

RESULT 9

```

US-08-967-506-19
; Sequence 19, Application US/08967506
; Patent No. 6096512
; GENERAL INFORMATION:
; APPLICANT: The Upjohn Company
; TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GalNAc-
; TITLE OF INVENTION: Polypeptide, N-Acetylglucosaminyltransferase
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pharmacia & Upjohn Company, Intellectual
; STREET: 301 Henrietta Street
; CITY: Kalamazoo
; STATE: Michigan
; COUNTRY: USA
; ZIP: 49001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/967,506
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/602,830
; FILING DATE: 13 No. 6096512ember 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Darnley Jr., James D.
; REGISTRATION NUMBER: 33,673
; REFERENCE/DOCKET NUMBER: 4755.P CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 616-833-2210
; TELEFAX: 616-833-8897
; TELEX: 224401
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 517 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-967-506-19

```

Query Match 34.0%; Score 115; DB 3; Length 517;

Best Local Similarity 43.6%; Pred. No. 8.3e-11;

Matches 230; Conservative 89; Mismatches 170; Indels 38; Gaps 16;

QY 88 GNGEGRP--YPMTDARVQAVRENGFNIVSDKSLNSLPDISHPNCNSKRYLETLP 145

```

Db 15 GPEMGKPVVLPKEDKMKEMKFNQFNLMASEMIALNRSLPDLVEGCKTKVYPDNL 74
QY 146 NTSIIIPFHEGSSSLRTVHSLVNSRPPELVAEITLVDFSDREHLKKPLJEDYM--ALFP 204
Db 75 TISVIVFHEANSTLRTVHSLVNSRPPHLEIIVLDASERDLKAPLESYVKKLKY 134
QY 205 SVRIIRTKKREGILRTTMLGASVATGVTITFLDSHCANVNMPLPDLIRIARNRKTIYCP 264
Db 135 PVAHVMEQSGSLIRARLGAASVKGQVITFDAAHECTVGMLEPLIARIKDRKTIYCP 194
QY 265 MIVIVDHDFFRYTQNG--DAMGAFDWEYTKIPIPLP---ELQKADPSDPFSPVMAAG 320
Db 195 IIDVIDDTPFEY--MAGSDMTYGFNMKINFRWYVPVOREMRKCDRTLPTVPTMAG 252
QY 321 LPAVDKRMFELGDPGRLGELMGEGEYBISFKVMGCGMMEIDPCSRVGHITYKYVPYKY 380
Db 253 LFSIDDDYFOEITGADMGIMGEMLEISFRIMQCGTLEIYTCSHVGHVFRKATPYTF 312
QY 381 PAGVS--LARNLRVAEVMWDEYAEYIYQRRPEYRHLSAGDVAVQKLRSSLNCKSPKMF 438
Db 313 PGGTGQIINKNNRRLAEVWMDDEFKNFFYIISPGVTKYDGDISSRLGLRHLQCRPFWSY 372
QY 439 MTKIAMD--LPKFPYPPVEPPAAAGWGEIRNVGTGLCADTKHGLGSPRLLEGVGRGEAA 496
Db 373 LENIYDPSQIPRHY-----FSLGEIRNVETWQCLDNNAKENEKXGIFNC--HGWG--- 421
QY 497 WNNMQUFTFTWRREDIRGDPQHTKKFCFPAISHTSPVTLYDCHSMKGNOLMYKRXDK--TL 555
Db 422 --GNQVFSYANKKEIRTD-----LCIDVSKLNGVPTMLKCHHLKGNOLMEYDVPVXLTL 473
QY 556 YHPSGSCMD-CSESDHRI-FMNTCNPSLTOQMLFEHTNSTVLEKF 600
Db 474 QHNSNQCLDKATDEDSQVPSIRDCS-GSRSQOMLNR--NVTLPEIF 517

```

RESULT 10

```

PCT-US94-02552-19
; Sequence 19, Application PC/TUS9402552
; GENERAL INFORMATION:
; APPLICANT: Elihammer, Ake P.
; TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GalNAc-
; TITLE OF INVENTION: Polypeptide, N-Acetylglucosaminyltransferase
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Upjohn Company, Corp. Intellectual
; STREET: 301 Henrietta Street
; CITY: Kalamazoo
; STATE: Michigan
; COUNTRY: USA
; ZIP: 49001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/02552
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Darnley Jr., James D.
; REGISTRATION NUMBER: 33,673
; REFERENCE/DOCKET NUMBER: 4755.P CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 616-385-5210
; TELEFAX: 616-385-6897
; TELEX: 224401
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 517 amino acids
; TYPE: amino acid

```


Db 520 TTFPHSGTCLSAVYRPEGRPDVQWRTCDALDKXQWISFE 559

RESULT 16

US-09-795-926-33

Sequence 33, Application US/09795926

Patent No. 6555669

GENERAL INFORMATION:

APPLICANT: Donoho, Gregory

APPLICANT: Hilbun, Erin

APPLICANT: Turner, C. Alexander Jr.

APPLICANT: Friedlich, Glenn

APPLICANT: Abuin, Alejandro

APPLICANT: Zambrowicz, Brian

APPLICANT: Sanders, Arthur T.

APPLICANT: Walke, D. Wade

APPLICANT: Wilganowski, Nathaniel L.

APPLICANT: Hu, Yi

APPLICANT: Kleke, James Alvin

APPLICANT: Potter, David George

TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND

TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME

FILE REFERENCE: LEX-0144-USA

CURRENT APPLICATION NUMBER: US/09/795,926

CURRENT FILING DATE: 2001-02-28

PRIOR APPLICATION NUMBER: US 60/185,920

PRIOR FILING DATE: 2000-02-29

PRIOR APPLICATION NUMBER: US 60/186,558

PRIOR FILING DATE: 2000-03-02

PRIOR APPLICATION NUMBER: US 60/191,849

PRIOR FILING DATE: 2000-03-24

NUMBER OF SEQ ID NOS: 47

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 33

LENGTH: 240

TYPE: PRT

ORGANISM: homo sapiens

US-09-795-926-33

Query Match 30.2%; Score 990; DB 4; Length 240;

Best Local Similarity 97.9%; Pred. No. 6.8e-98;

Matches 189; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 1 MRKKEKRLQAVLVLAALVLPVGLMALYRERQDPGTGSGAAYAPAAQSGSHSRK 60

Db 1 MRKKEKRLQAVLVLAALVLPVGLMALYRERQDPGTGSGAAYAPAAQSGSHSRK 60

QY 61 KTFPLGDGQKLDKMDHKEAIRDAQRVNGEGGRPYMTDAERVDQAYRENGFNIVYSDK 120

Db 61 KTFPLGDGQKLDKMDHKEAIRDAQRVNGEGGRPYMTDAERVDQAYRENGFNIVYSDK 120

QY 121 ISLNRSIPDIRHNCNCKRYLETLPNTSIIIPFHNCGSSLLRTVHSVLRSPPELVAEI 180

Db 121 ISLNRSIPDIRHNCNCKRYLETLPNTSIIIPFHNCGSSLLRTVHSVLRSPPELVAEI 180

QY 181 VLVDFFSDRHLK 193

Db 181 VLVDFFSDRGRIR 193

QY 181 VLVDFFSDRGRIR 193

Db 181 VLVDFFSDRGRIR 193

QY 181 VLVDFFSDRGRIR 193

Db 181 VLVDFFSDRGRIR 193

QY 181 VLVDFFSDRGRIR 193

Db 181 VLVDFFSDRGRIR 193

QY 181 VLVDFFSDRGRIR 193

Db 181 VLVDFFSDRGRIR 193

QY 181 VLVDFFSDRGRIR 193

Db 181 VLVDFFSDRGRIR 193

QY 181 VLVDFFSDRGRIR 193

Db 181 VLVDFFSDRGRIR 193

QY 181 VLVDFFSDRGRIR 193

Db 181 VLVDFFSDRGRIR 193

QY 181 VLVDFFSDRGRIR 193

Db 181 VLVDFFSDRGRIR 193

QY 181 VLVDFFSDRGRIR 193

Db 181 VLVDFFSDRGRIR 193

QY 181 VLVDFFSDRGRIR 193

Db 181 VLVDFFSDRGRIR 193

QY 181 VLVDFFSDRGRIR 193

Db 181 VLVDFFSDRGRIR 193

APPLICANT: Walke, D. Wade

APPLICANT: Wilganowski, Nathaniel L.

APPLICANT: Hu, Yi

APPLICANT: Kleke, James Alvin

APPLICANT: Potter, David George

TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND

TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME

FILE REFERENCE: LEX-0144-USA

CURRENT APPLICATION NUMBER: US/09/795,926

CURRENT FILING DATE: 2001-02-28

PRIOR APPLICATION NUMBER: US 60/185,920

PRIOR FILING DATE: 2000-02-29

PRIOR APPLICATION NUMBER: US 60/186,558

PRIOR FILING DATE: 2000-03-02

PRIOR APPLICATION NUMBER: US 60/191,849

PRIOR FILING DATE: 2000-03-24

NUMBER OF SEQ ID NOS: 47

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 37

LENGTH: 209

TYPE: PRT

ORGANISM: homo sapiens

US-09-795-926-37

Query Match 30.1%; Score 988; DB 4; Length 209;

Best Local Similarity 100.0%; Pred. No. 8.8e-98;

Matches 189; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MRKKEKRLQAVLVLAALVLPVGLMALYRERQDPGTGSGAAYAPAAQSGSHSRK 60

Db 1 MRKKEKRLQAVLVLAALVLPVGLMALYRERQDPGTGSGAAYAPAAQSGSHSRK 60

QY 61 KTFPLGDGQKLDKMDHKEAIRDAQRVNGEGGRPYMTDAERVDQAYRENGFNIVYSDK 120

Db 61 KTFPLGDGQKLDKMDHKEAIRDAQRVNGEGGRPYMTDAERVDQAYRENGFNIVYSDK 120

QY 121 ISLNRSIPDIRHNCNCKRYLETLPNTSIIIPFHNCGSSLLRTVHSVLRSPPELVAEI 180

Db 121 ISLNRSIPDIRHNCNCKRYLETLPNTSIIIPFHNCGSSLLRTVHSVLRSPPELVAEI 180

QY 181 VLVDFFSDR 189

Db 181 VLVDFFSDR 189

QY 181 VLVDFFSDR 189

Db 181 VLVDFFSDR 189

QY 181 VLVDFFSDR 189

Db 181 VLVDFFSDR 189

QY 181 VLVDFFSDR 189

Db 181 VLVDFFSDR 189

QY 181 VLVDFFSDR 189

Db 181 VLVDFFSDR 189

QY 181 VLVDFFSDR 189

Db 181 VLVDFFSDR 189

QY 181 VLVDFFSDR 189

Db 181 VLVDFFSDR 189

QY 181 VLVDFFSDR 189

Db 181 VLVDFFSDR 189

QY 181 VLVDFFSDR 189

Db 181 VLVDFFSDR 189

QY 181 VLVDFFSDR 189

Db 181 VLVDFFSDR 189

QY 181 VLVDFFSDR 189

Db 181 VLVDFFSDR 189

QY 181 VLVDFFSDR 189

Db 181 VLVDFFSDR 189

QY 181 VLVDFFSDR 189

Db 181 VLVDFFSDR 189

QY 181 VLVDFFSDR 189

Db 181 VLVDFFSDR 189

QY 181 VLVDFFSDR 189

Db 181 VLVDFFSDR 189

QY 181 VLVDFFSDR 189

Db 181 VLVDFFSDR 189

QY 181 VLVDFFSDR 189

Db 181 VLVDFFSDR 189

QY 181 VLVDFFSDR 189

Db 181 VLVDFFSDR 189

QY 181 VLVDFFSDR 189

Db 181 VLVDFFSDR 189

NUMBER OF SEQ ID NOS: 550
 SEQ ID NO 196
 LENGTH: 552
 TYPE: PRT
 ORGANISM: Homo Sapien
 US-10-140-002-196

Query Match 29.6%; Score 971; DB 4; Length 552;

Best Local Similarity 37.2%; Pred. No. 3,1e-95; Indels 98; Gaps 19;
 Matches 223; Conservative 90; Mismatches 199;

QY 1 MRREKELLQAVLVLAALVLLPVGVMALYRERQ-PDG---TPGGGAVALPAAGQS 55
 1 MRRLTRLVLPVGVLTITVLL---FFWVTRKLEVTGPEVQPKRSDA----- 47
 DB 56 HSROKTFPLGDDQCKLQMDH-----KEAIRDAORVNGEGRPYPMTDARVQAY 108
 48 -----DMDLMDQFDERRYLNKKRVG-----DDPY 74
 QY 109 RENGFIYVSDKLSLNSLPDIRHPNCSKRYLETLPNTSIIIPFNHSGMSLLRTVHSV 168
 75 KLVAFNORSESRISNRAIPDTRHLRCTLVYCTDLPPTSIITFNHANSSTLKTIRSV 134
 QY 169 LNSRPPELVAEIYLVDFDS-DREHLKKPLEDYALPFSVRILRTKKEGLIRTMGLASV 227
 135 LNRPTHLIREIILVDFDSNDPDDCKLIK-----LPRVKCLRNNEGGLVRSIRGADI 189
 QY 228 ATGCVITFLDSHCANVMLEPLLDIRANRKTIVCPMIDIVDHDFFRYETQAGDANRGA 287
 190 AOGTTFLDSHCENRDMQLPLHRYVEDYTRVVCVIDIINDFTYIESASE-LRGG 248
 QY 288 FDMEMYKRIPIPELOKA---DPSDFESPVMAGLFAVDKRMFELGVDPELEIIMG 344
 249 FDSLHQWEOLEPE-QARRLDPEIRPTIINGGLFVIDKAFDYLKIDMDMDIWGG 307
 QY 345 EOYISFKVMCGRMEDIPCSRIGHIYRKIVPYKPVAG--VSLARLKVAEVWMDYA 402
 308 ENFISIRVMCGSLIIVPCSRGIVFRKHPVFPDGNANTYIKTKTAAEVMDEYK 367
 QY 403 EYIYORPEYRHLASGVAVOKKLRSSLNCSFFWMTKIAMDLPKFYPPVEPP--AAA 459
 368 QYVVAARPFALERPFGVNBSRLDKRLRCQSPFKVYLENI-----YPELSIPKESSIQ 420
 DB 460 WGEIRNVTGICADTKGALGSLRLGECVRGGEAAMNNQVFTFTREDIRGDDQHT 519
 421 KGNROROKCLESORNNQETPNLKLSPCAVKGEDA--KSQVAAFTYTOIL-----Q 472
 QY 520 KKFCFDAIS--HTSPVTLVDCHSMKGNOLMKYRKDKTLVHPVSGSCMDC-----SSDH 571
 473 EELCLAVITLFPKAPVVLVLCNKGGDDROQT-KTGSHEIASHLCLDTDMFGDTENGK 531
 QY 572 RIFMNTCPSSLIQQW 587
 DB 532 EIVNPPCESSIIMSQHW 547

RESULT 19

US-09-270-767-33733
 Sequence 33733, Application US/09270767
 Patent No. 6703491
 GENERAL INFORMATION:
 APPLICANT: Homburger et al.
 TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 FILE REFERENCE: File Reference: 7326-094
 CURRENT APPLICATION NUMBER: US/09/270,767
 NUMBER OF SEQ ID NOS: 62517
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 33733
 LENGTH: 302
 TYPE: PRT
 ORGANISM: Drosophila melanogaster
 US-09-270-767-33733

Query Match 28.7%; Score 941; DB 4; Length 302;
 Best Local Similarity 59.7%; Pred. No. 1.9e-92; Indels 6; Gaps 4;
 Matches 178; Conservative 44; Mismatches 70;

QY 80 IRDAORVNGEGRPYPMTDARVQAYR---ENGFNIVSDKLSLNSLPDIRHPNCS 136
 2 MENDARVVGEGGKASTLDDSEGRLEKMSLENGFNLLSDISIVNSRVPDIRHPLCR 61
 QY 137 SKRYLETPTSIIIPFNHSGMSLLRTVHSLNRPPELVAEIYLVDFSDSEHLKKPL 196
 62 KKEVYAKLPVSVYIIIFVNEYSVLNRSVSLNRPPELMKIIIVDHSDERYLQKEL 121
 QY 197 EDMA-LFPEVRILRTKKEGLIRTMGLASVATGVITFLDSHCANVMLEPLLDIRA 255
 122 EYIAHEFKVVRVRLPRRTGLGAPAGARNAETVLFDSHVEANVMLEPLLDIRA 181
 QY 256 RNRKIVCPMIDIVDHDFFRYETQAGDANRGAFFDMYKRIPIPELOKADSDPESP 315
 182 LNRRTAVCPIDVIDHTNFIYRAQ--DEGARGADEFFYKRLPLPEDLK-HPADPFKSP 239
 QY 316 WAGGLFAVDKRMFELGVDPELEIIMGGEQYISFKVMCGRMEDIPCSRIGHIYR 373
 240 IMAGGLFAISKEPFWEIYDEGLDIWGEQYELSKRIWCGEMVDAPCSRIGHIYR 297

RESULT 20

US-09-376-856-2
 Sequence 2, Application US/09376856
 Patent No. 6235510
 GENERAL INFORMATION:
 APPLICANT: KIMAR, SANJAY
 APPLICANT: VAN HORN, MARION M.
 APPLICANT: LARK, MICHAEL
 TITLE OF INVENTION: ppgAntase-76
 FILE REFERENCE: GP-70638
 CURRENT APPLICATION NUMBER: US/09/376,856
 CURRENT FILING DATE: 1999-08-18
 NUMBER OF SEQ ID NOS: 2
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 2
 LENGTH: 639
 TYPE: PRT
 ORGANISM: HOMO SAPIENS
 US-09-376-856-2

Query Match 28.2%; Score 926; DB 3; Length 639;
 Best Local Similarity 35.5%; Pred. No. 2.8e-90; Indels 80; Gaps 17;
 Matches 216; Conservative 96; Mismatches 216;

QY 4 KERKLLQAVLVLAALVLLPVGMLALYRERQDGTTPGGGAVALPAAGGSHSRQKTF 63
 91 REDQLVAVLPPA-----RNNQOGRRGGSYRLIKOPRRQ----- 126
 QY 64 FLDDGQCKLQMDHKEAIRDAORVNGEGRPYPMT---DAERVOAYENGFNIVSD 119
 127 -----DKEAPKRWGADDEGEVSESELIPFSLDPGLQEA-----LSA 165
 QY 120 KLSLNSLPDIRHPNCSKRYLETLPNTSIIIPFNHSGMSLLRTVHSLNRPPELVAE 179
 166 RIFLQRALPEVRHPLCQHPQDSLPASVILCFHDEAMGTLRTVHSLIDTVPRALKE 225
 DB 180 IYLVDFSDREHLKKPLEDYALPFSVRILRTKKEGLIRTMGLASVATGVITFLDSH 239
 226 IIVVDLSQGGKLSLSEVVARLEGGKLRNSKRGALRAARMLGATRAVDLVFWDAAH 285
 QY 240 CEANVMLEPLLDIRANRKTIVCPMIDIVDHDFFRYETQAGDANRGAFFDMYKRIPI 299
 286 CECHPGWLEPLSLRIGDRSVSPYVIDDMKTFQY-YRSKLOQGVADLDFHKEPL 344
 DB 300 PPELOKA--DPSDFESPVMAGGLFAVDKRMFELGVDPELEIIMGGEQYISFKVMCG 357
 345 PEHVRKALQSPISPIPSVVGSEVAMDRHYFQNTGAYDLSLARGENLELSFKAWLGC 404

QY 358 GRMEDIPCSRVGHITRYKTVYKVPAGVSLARNLKRVAEVMWDEAEYIYORREYRLISA 417
 DB 405 GSVEILPCSRVGHITRYQODSHSLPDQDQATLRNRVRIAEIETWLGSPKEITFYHGSPEAFSLSK 464
 QY 418 G---DVAVQKRLSSLNCKSPKFMWTIAMDLPEFYPPVEPPAAWGEIRNVGTGLCAD- 473
 DB 465 AEKPDCEMRLOQLORLGCRTFHFHFLANV---YDELTPS-EPREPSFGKLNHTGLGACD 520
 QY 474 -THGALGSPRLREGCVRGGEAAMNMVFTFTWRBDRPDQHTKFCFPDAISHTSP 532
 DB 521 QAEGDILGCEMVLAPCDSR-----QQOYLQHTSRKEIHFGSPOR---LCFAV---ROEQ 569
 QY 533 VTLVDCSHMKG---NQLMKYRKDKTLVHPVSGSCMDC--SESDBRIFMNTCNPSLSLTOQ 586
 DB 570 VILQNC-TEEGALAHQOHMDFOENGMIIVHILSGKCEAVVQENNKDLYLRPCD-GKARQO 627
 QY 587 WLEHTNS 594
 DB 628 WRFQJINA 635

RESULT 21

US-09-347-488-2
 ; Sequence 2, Application US/09347488
 ; Patent No. 6239266
 ; GENERAL INFORMATION:
 ; APPLICANT: Munroe, David
 ; APPLICANT: Pribill, Ingrid
 ; TITLE OF INVENTION: ZAP-3 TUMOR ASSOCIATED GENES AND THEIR
 ; FILE OF INVENTION: USES
 ; FILE REFERENCE: GEN-2PRV
 ; CURRENT APPLICATION NUMBER: US/09/347,488
 ; CURRENT FILING DATE: 1999-07-02
 ; NUMBER OF SEQ. ID NOS: 2
 ; SOFTWARE: FASTSEQ for Windows Version 3.0
 ; SEQ ID NO 2
 ; LENGTH: 638
 ; TYPE: PRF
 ; ORGANISM: H. sapiens
 ; US-09-347-488-2

Query Match 28.1%; Score 920; DB 3; Length 638;
 Best Local Similarity 35.4%; Pred. No. 1,2e-89;
 Matches 215; Conservative 96; Mismatches 217; Indels 80; Gaps 17;

QY 4 KERKILQAVLVLAALVILPVGIMALYREROPDTRGSGGAANAAPAGGSGSRCKTF 63
 DB 91 REDQLVAVALPQA-----KRNQSGRRGSTRILIKOPRQ----- 126
 QY 64 FLGDGQKLKDMHDKAIRDAORVNGEOGRPYPMT----DAERVDQYRENGFNIVSD 119
 DB 127 -----DKEAPKRDWGADEDEVEEBELTFPSLDPRGLQA-----LSA 165
 QY 120 KISLNSLPDIRHPNCKSKYLETLPNTSIIIPPHNGSSLLRTVSHVLSRSPPELVAE 179
 DB 166 RLPQRLAPYVRHPLCQCHPODSLPTASVILCFHDEAWSTLLRTVSHIIDTVPRALKE 225
 QY 180 IYLVDDFSDBREHLKKPLEDMALPFSVRLITKKBELITRMGASAVANGDVTTFDISH 239
 DB 226 IILVDDLSQGOQLSALSERVAREGLVLRNSRRLSAIRARMGATRAATGVALVFEVDH 285
 QY 240 CEANVNMVLPPLDIRIARNRTIYCPMIDVIDHDQFRYETOAGDAMRGAPFMEWYKRIPI 299
 DB 286 CECHPGWLEPLSLIAGDRSRVVSPIVDIMKTFQY-YESKDLQKRVLDKDLFHWEP 344
 DB 345 PEHVRKALQSPISPIRSFVVGGEVAVADRHVFQNTGAYDSLMSLRGENELSLFKAWLCG 404
 QY 300 PPELOKA--DPSDFESPVWAGGLFAVDKRMFWELGSDPGLIWSGEQYISFKWMC 357
 DB 358 GRMEDIPCSRVGHITRYKTVYKVPAGVSLARNLKRVAEVMWDEAEYIYORREYRLISA 417
 DB 405 GSVEILPCSRVGHITRYQODSHSLPDQDQATLRNRVRIAEIETWLGSPKEITFYHGSPEAFSLSK 464

QY 418 G---DVAVQKRLSSLNCKSPKFMWTIAMDLPEFYPPVEPPAAWGEIRNVGTGLCAD- 473
 DB 465 AEKPDCEMRLOQLORLGCRTFHFHFLANV---YDELTPS-EPREPSFGKLNHTGLGACD 520
 QY 474 -THGALGSPRLREGCVRGGEAAMNMVFTFTWRBDRPDQHTKFCFPDAISHTSP 532
 DB 521 QAEGDILGCEMVLAPCDSR-----QQOYLQHTSRKEIHFGSPOR---LCFAV---ROEQ 569
 QY 533 VTLVDCSHMKG---NQLMKYRKDKTLVHPVSGSCMDC--SESDBRIFMNTCNPSLSLTOQ 586
 DB 570 VILQNC-TEEGALAHQOHMDFOENGMIIVHILSGKCEAVVQENNKDLYLRPCD-GKARQO 627
 QY 587 WLEHTNS 594
 DB 628 WRFQJINA 635

RESULT 22

US-08-648-298-2
 ; Sequence 2, Application US/08648298
 ; Patent No. 5871990
 ; GENERAL INFORMATION:
 ; APPLICANT: Henrik Clausen
 ; APPLICANT: Eric Paul Bennett
 ; TITLE OF INVENTION: UDP-N-acetyl-alpha-D-galactosamine:polypeptide
 ; TITLE OF INVENTION: N-acetyl galactosaminyltransferase GalNAc-T3
 ; NUMBER OF SEQUENCES: 19
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Darby & Darby PC
 ; STREET: 805 Third Avenue
 ; CITY: New York
 ; STATE: NY
 ; ZIP: 10022
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/648,298
 ; FILING DATE: 15-JUN-1996
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Green, Reza
 ; REGISTRATION NUMBER: 38,475
 ; REFERENCE/DOCKET NUMBER: 4035/05865
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212527700
 ; TELEFAX: 2127536237
 ; TELEX: 236687
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 633 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; ORIGINAL SOURCE:
 ; ORGANISM: Homo sapiens
 ; TISSUE TYPE: Submaxillary gland
 ; US-08-648-298-2

Query Match 27.9%; Score 916; DB 2; Length 633;
 Best Local Similarity 37.6%; Pred. No. 3,3e-89;

Matches 202; Conservative 100; Mismatches 177; Indels 58; Gaps 18;
 QY 90 GEGGRPYMTDAERVDQAVRENG-----ENIYVSDKISLNRSL-PDIRHPNCKSKRY--L 141
 DB 122 GASGKAFKTTNLSVEQKKEGAEACGCFNAFASDRISLHRLDGPTRPECEIQKFKRC 181
 QY 142 ETLPTNSIIIPPHNEWSSLLRTVSHVLSRSPPELVAEIVYDDFSDBREHLKKPLEDMYA 201
 DB 182 PPLPTTSVILVFNENWSTLLRTVSHVLSRSPPELVAEIVYDDFSDBREHLKKPLEDMYA 241

QY 202 LPSPVRLTRTKREGGLIRTRMLGASVATGVTITFLDSHCSEANVMMLPPLDRIARANKTI 261
 DB 242 QFSTVKIVRQREKGLITLRLGKTVATETITFLDHCCEFCYGMLEPLARIENYAV 301
 QY 262 VCPMIDVIDHDDFRYETQA---GDAMGAFDMEWY---KRIPIPELOKADPSDPSPSV 316
 DB 302 VSPDIASIDINTFENKPSPYGSNNHNGNFMWSLSFGMESLPDHEKQRKDETYPIKTP 361
 QY 317 MAGGLFAVDKRMFELGCGYDGLIEINGEYIEISFKYVMCGGRNEDIPCSRVGHIYKTY 376
 DB 362 FAGGLFISKEYFEYISYDEMEIEMGENIEMSFVWCGGLEIMPCSVVGHVFRSKS 421
 QY 377 PYKYPAGVS-LARLKKVAEVMDEYAEYIYORPE---YRHLASGVAOKKLSLSLN 431
 DB 422 PHSPFKGTQYIARQVIAEVMDEYKEIFRRTTDAKIVKQAFGLSKREFFIKRLR 481
 QY 432 CKSEKFMETKIAMDLPKFYPPVEPPAAWGEIRNVGTGLCADT-KHGLASPLRLBECVR 490
 DB 482 CKNTWTYLNMI---YPEVYVDLNPVLS-GYIKSVGQPLCLDVENNOGKPLIMYTC-H 536
 QY 491 GREGAANNNOVFTFTREDIRPDPOHT--KKECFDAI-----SHTSPVTLTY 536
 DB 537 GLG-----GNQYFYSQAQHEIR-----HNTQKELCLHAAQGLVOLKACTYKGHKTIVT-- 584
 QY 537 DCHSMKGNOLMKYRKDTLVHPVSGCMDCSESDHRIFMNTCNPSLSLTOCKLFEHTN 593
 DB 585 -----GBQIWEIOKDLINPLMKCLS-ANGHPSLV-SCNPSDPLQKWLISOND 633

RESULT 23

US-09-795-926-23
 Sequence 23, Application US/09795926
 Patent No. 6553669
 GENERAL INFORMATION:
 APPLICANT: Donoho, Gregory
 APPLICANT: Hilbun, Erin
 APPLICANT: Turner, C. Alexander Jr.
 APPLICANT: Friedrich, Glenn
 APPLICANT: Abidin, Alejandro
 APPLICANT: Zambrowicz, Brian
 APPLICANT: Sands, Arthur T.
 APPLICANT: Walke, D. Wade
 APPLICANT: Wilganowski, Nathaniel L.
 APPLICANT: Hu, Yi
 APPLICANT: Kieke, James Alvin
 APPLICANT: Poter, David George
 TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
 FILE REFERENCE: LEX-0144-USA
 CURRENT APPLICATION NUMBER: US/09/795, 926
 CURRENT FILING DATE: 2001-02-28
 PRIOR APPLICATION NUMBER: US 60/185,920
 PRIOR FILING DATE: 2000-02-29
 PRIOR APPLICATION NUMBER: US 60/186,558
 PRIOR FILING DATE: 2000-03-02
 PRIOR APPLICATION NUMBER: US 60/191,849
 NUMBER OF SEQ ID NOS: 47
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 23
 LENGTH: 224
 TYPE: PRT
 ORGANISM: homo sapiens
 US-09-795-926-23

Query Match 24.5%; Score 804; DB 4; Length 224;
 Best Local Similarity 98.7%; Pred. No. 6.7e-76;
 Matches 154; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 98 MTDARVDOAVRENGFNIVYSKISLNSLPDIRHPNCSKRYLETLPNTSIIIPFNHG 157
 DB 1 MTDARVDOAVRENGFNIVYSKISLNSLPDIRHPNCSKRYLETLPNTSIIIPFNHG 60

QY 158 WSSLTRVSHVLSRSPPELVAEIVLVDDFSDREHLKKPLEDYALPEPSVRLTRTKREG 217
 DB 61 WSSLTRVSHVLSRSPPELVAEIVLVDDFSDREHLKKPLEDYALPEPSVRLTRTKREG 120
 QY 218 IRTTMLGASVATGVTITFLDSHCSEANVMMLPPLDRIARANKTI 261
 DB 121 IRTTMLGASVATGVTITFLDSHCSEANVMMLPPLDRIARANKTI 261

RESULT 24

US-09-973-457-2
 Sequence 2, Application US/0973457
 Patent No. 6703230
 GENERAL INFORMATION:
 APPLICANT: Kapeller, Libermann, Rosana
 TITLE OF INVENTION: 471/4, A NOVEL HUMAN GLYCOSYLTRANSFERASE
 FILE REFERENCE: 10448-099001
 CURRENT APPLICATION NUMBER: US/09/973,457
 CURRENT FILING DATE: 2001-10-09
 PRIOR APPLICATION NUMBER: 60/238,849
 PRIOR FILING DATE: 2000-10-06
 NUMBER OF SEQ ID NOS: 6
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 2
 LENGTH: 603
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-973-457-2

Query Match 22.7%; Score 744; DB 4; Length 603;
 Best Local Similarity 32.3%; Pred. No. 1e-70;
 Matches 202; Conservative 86; Mismatches 230; Indels 108; Gaps 23;

QY 3 RKEKRLLOAVATVLAALVLLPNTGLMALYEROPDGTGGGGAAPAG-QGSHSROK 61
 DB 5 RKIRTLTVNIIIVFGIV-----LFSVGRLO-----GRSGLVNIIVSGDRVRSHAK 53
 QY 62 TFFLGQGGKLDN-HDKAIRDQORV-----GNGEQRPYV-TDAERYDQA 107
 DB 54 VGTLDGRREALQRLDHEEVVYQNLGAKPIGLVEGPGIGGGLAATLRDGOEABGK 113
 QY 108 YRNGFNIVYSDKISLNSLPDIRHPNCSKRYLETLPNTSIIIPFNHGSSILRTVHS 167
 DB 114 YEEYVNAQUSDISLDRSIPDYPRKCRWMSYAQDPQVSVFIFVNEALSVILSVHS 173
 QY 168 VLNRSPPELVAEIVLVDDFSDREHLKKPLEDYV-ALFPS-VRIARTKREGGLIRTRMLGA 225
 DB 174 VVHNTBSQLKEYIYLVDDNSDNTLAKFNLDQYNNKRYPGIVKTVRSRRREGGLIRARLQGW 233
 QY 226 SVATGVTITFLDSHCSEANVMMLPPLDRIARANKTIYCPMIDVIDHDDFRYETQAGDAMR 285
 DB 234 KAATAVAVGPFDAHVEFNTGMAPALSRIRREDRRITVPAIDNIKYSTFEVQ-QYANAAN 292
 QY 286 GADMEWYTKRIPIPE--LQKADPSDPSPVMAAGLPAVDKMFELGCGYDGLIEING 343
 DB 293 G-YNMGIMCMYI-IPQDMIDRGDESAPIRTPAMIGCSFVADREYFGDIGLDPGMEVYG 350
 QY 344 GEYIEISFKYVMCGGRNEDIPCSRVGHIYKVPYKYPAGVSLARNIKRVAEVMDEYAE 403
 DB 351 GENVELAMRVWCGGSGSEVUPCSRVAHIEKTRPYNNDDIYAKRALRAAEVMDDFKS 410
 QY 404 YIYQ--RPEYHLSAGDAVQKJLSSLNCKSFKWEMTKIAMDLPKFYPP--VEPPAA 458
 DB 411 HYYMANNIPMSNPGVDGVSERIALRQRLKCRSFKYLENV-----YPERKVVNNTL 463
 QY 459 AMGEIRV-VGTGLCADTKHGLASPLRLBECVARGGAANNNOVFTFTREDIRPDPO 517
 DB 464 TYGVRNSKASAVCLD--QGA-----ED--GD-- 486
 QY 518 HTKFCFCDALSHSPVTLVYDCHSMKGNOLMKYRKD-----KTLVHPVSGCMDCSE 568


```

; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSPERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/186,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 143
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-795-926-21

Query Match      14.7%; Score 483; DB 4; Length 143;
Best Local Similarity 95.8%; Pred. No. 1.2e-43;
Matches 92; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy      98 MTDARVDAQYRENGFNIVYSDKISLNRSLPDIRHPNCSKRYLETLPNTSIIPFNEG 157
Db      1 MTDARVDAQYRENGFNIVYSDKISLNRSLPDIRHPNCSKRYLETLPNTSIIPFNEG 60

Qy      158 WSSLRTVSHSVLNRSPELVAVIYLVDDFSDRGIR 193
Db      61 WSSLRTVSHSVLNRSPELVAVIYLVDDFSDRGIR 96

RESULT 29
US-09-795-926-25
; Sequence 25, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedlich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSPERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 112
; TYPE: PRT
; ORGANISM: homo sapiens
```

```

US-09-795-926-25

Query Match      14.7%; Score 481; DB 4; Length 112;
Best Local Similarity 100.0%; Pred. No. 1.4e-43;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      98 MTDARVDAQYRENGFNIVYSDKISLNRSLPDIRHPNCSKRYLETLPNTSIIPFNEG 157
Db      1 MTDARVDAQYRENGFNIVYSDKISLNRSLPDIRHPNCSKRYLETLPNTSIIPFNEG 60

Qy      158 WSSLRTVSHSVLNRSPELVAVIYLVDDFSDR 189
Db      61 WSSLRTVSHSVLNRSPELVAVIYLVDDFSDR 92

RESULT 30
US-09-973-457-6
; Sequence 6, Application US/09973457
; Patent No. 6703230
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 47174, A NOVEL HUMAN GLYCOSYLTRANSFERASE
; FILE REFERENCE: 10448-099001
; CURRENT APPLICATION NUMBER: US/09/973,457
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/238,849
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus sequence
US-09-973-457-6

Query Match      13.7%; Score 450.5; DB 4; Length 149;
Best Local Similarity 53.1%; Pred. No. 4.3e-40;
Matches 78; Conservative 28; Mismatches 36; Indels 5; Gaps 2;

Qy      305 KADSPDPFESPVNAGGLFAVDRKFWELGGYDGLIHWGGEQYRISPKVMCGRMDDIP 364
Db      1 KADPTDPRSPVNAAGGLFAVDRKFWELGGYDGLIHWGGEQYRISPKVMCGRMDDIP 60

Qy      365 CSRVGHIYRKYVPYKVPAGVS---LARNLKRVAEVMDBYAEYIYQRRPEYHL--SAGD 419
Db      61 CSRVGHIYRKYVPYKVPAGVS---LARNLKRVAEVMDBYAEYIYQRRPEYHL--SAGD 120

Qy      420 VAVQKIRSSLNCKSPKFWNTKIAMDI 446
Db      121 ISERKEIREKLQCKSPKFWNTKIAMDI 147

RESULT 31
US-09-270-767-58910
; Sequence 58910, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 58910
; LENGTH: 161
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-58910
```

Query Match 13.1%; Score 428.5; DB 4; Length 161;
 Best Local Similarity 54.5%; Pred. No. 1.1e-37;
 Matches 84; Conservative 2; Mismatches 45; Indels 1; Gaps 1;

QY 129 DIHPNCSKRYETLPNTSIIIPHNCGSSLIPTVSHVSNRSPPEL 168
 DB 3 DTLEERHMDYPPDIPRTSVIIFHNCGSVLMRTVSHVSDSPMTMHEILLVDFSD 62

QY 189 REHLKKLEEDYMLPFS-VAILRTKKREGILRTRMGASVATGDTITFLDSHCEANNWL 247
 DB 63 KENLRSQLDEIVLOFGKLVIVIRNKEREGILRTSRAMEATGIVIFLDAHCEVNTNWL 122

QY 248 PPLLDRIARNKRTIVCEPMIDVIDHDDEFRYETOG 281
 DB 123 PPLAPIYRDRTWTVPITIDGIDHKNEFRPVYG 156

RESULT 32
 US-09-513-999C-7570
 ; Sequence 7570, Application US/09513999C
 ; Patent No. 6783961
 ; GENERAL INFORMATION:
 ; APPLICANT: Dumas Milne Edwards, J. B.
 ; APPLICANT: Duclert, A.
 ; APPLICANT: Giordano, J. Y.
 ; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
 ; Patent No. 6783961
 ; FILE REFERENCE: 59.US2.REG
 ; CURRENT APPLICATION NUMBER: US/09/513.999C
 ; CURRENT FILING DATE: 2000-02-24
 ; PRIOR APPLICATION NUMBER: US 60/122,487
 ; PRIOR FILING DATE: 1999-02-26
 ; NUMBER OF SEQ ID NOS: 36681
 ; SOFTWARE: Patent.pm
 ; SEQ ID NO 7570
 ; LENGTH: 122
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-513-999C-7570

Query Match 8.8%; Score 287; DB 4; Length 122;
 Best Local Similarity 47.1%; Pred. No. 1.3e-22;
 Matches 56; Conservative 29; Mismatches 32; Indels 2; Gaps 1;

QY 117 VSDKISLRSLPIRHNCSKRYLETLPNTSIIIPHNCGSSLIPTVSHVSNRSPPEL 176
 DB 3 ISRLGYHRDVPDIPRNACKEKFPDLPASAVVICYNFAFSLLRTVSHVIDRTPAHL 62

QY 177 VAEIVLVDFSDREHLKKPLEDYMLF-PSVRLIRTKKREGILRTRMGASVATGDVI 233
 DB 63 LHEIILLVDDSDPDDKGLDEIVQKYLPGKIKIVIRTKREGILRGMIGAAHATGBVL 121

RESULT 33
 US-09-270-767-32872
 ; Sequence 32872, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 32872
 ; LENGTH: 148
 ; TYPE: PRT
 ; ORGANISM: Drosophila melanogaster
 US-09-270-767-32872

Query Match 8.5%; Score 277.5; DB 4; Length 148;
 Best Local Similarity 37.5%; Pred. No. 1.9e-21;

Matches 63; Conservative 23; Mismatches 41; Indels 41; Gaps 7;

QY 388 RNLRVAEYWMDEYAEIYQRRPE-YRHLSDGAVAVQKLSLNCKSFMTKIAMDL 446
 DB 6 KNTKRAEAWWWEYKLVYSHGDGLYESVDPDGLTEQKIRTKLNCKSRKMFEEVAFDL 65

QY 447 PKFYPVEPPAAMGEIRVGT-GLCADT-----KHGALG-----SPRLREGVGR 492
 DB 66 MKTYPVDEPPSYVMGALQVGVGNQLCLDTLGRKGNKMGVYACADIKTPOR----- 117

QY 493 GEAAANNQVFTTREDIRPDGPQHTKFCFPAISHTSPVLYDCHS 540
 DB 118 -----TQFWELSWRDLR-----LRKKECLD-----VOIMDNA 147

RESULT 34
 US-09-270-767-46209
 ; Sequence 46209, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 46209
 ; LENGTH: 313
 ; TYPE: PRT
 ; ORGANISM: Drosophila melanogaster
 US-09-270-767-46209

Query Match 7.4%; Score 241.5; DB 4; Length 313;
 Best Local Similarity 27.5%; Pred. No. 5e-17;
 Matches 86; Conservative 40; Mismatches 112; Indels 75; Gaps 15;

QY 345 EGYEISFKYWMCGRMEDIPCSRVGHYRKYPYKVPACVS--LAEHLKRVAEVWMDXA 402
 DB 1 ENVEMSFRIWCGGRVIEPCHSVGHVFRSSTPTFGGMSVLTNNLRATVMWDM- 59

QY 403 EYIYQRRPEYRLSAGD--VAVQKLSLNCKSFMTKIAMDLPKFYPVEPPA 459
 DB 60 QYFIMLYTSGLTGAKDKVNTVERVALRERLQCKPSPVLENI-W--PEHFFPA--PDRF 114

QY 460 WGEIRV-ETGLCA-----DPRGALSGPRL-GR 490
 DB 115 FGKIWLDETECAQVSKRMKLPGRALSREWKAFEEIDSKAEELMALIDLRDKLR 174

QY 491 -----GRGEAANNQVFTTREDIRPDGP-----QHTK-----KCFD 525
 DB 175 PLKEDYPRSSLSAVYTDCTSHAGSMDMFTFKGIMINDVCLTYRQOKGVIGMLKN 234

QY 526 AISHTSPVTLTYDCHSNKQNLWKYRD-KTYLHPVSGSCMDCSESDR-----IFMN 576
 DB 235 RNATTSNWMIAQAS-DSSQLWTYDMDTQOISHRDTKLCLTKAATNSRLQVEKVLAM 293

QY 577 TCNPSSLTQOMLF 589
 DB 294 ECDPXDITQKMGF 306

RESULT 35
 US-09-270-767-43809
 ; Sequence 43809, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 43809
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-43809

Query Match 7.2%; Score 234.5; DB 4; Length 300;
Best Local Similarity 30.0%; Pred. No. 2.6e-16;
Matches 63; Conservative 33; Mismatches 97; Indels 17; Gaps 6;

QY 383 GVSANLAKRVAWMD-VAEYTORPEYRHLSAGDVAVOKKLSLNCSEPKMFMTK 441
DB 32 GPITTYNKKVITWFDTHKEFYTRPLARLDNGDISQALAKKRLKSKSFQWMDH 91
QY 442 IAWDLKPEYPPVPPAAMGEIRNVGTGLCADTKHGLSGPLRLGCGVRGGEAAMNMQ 501
DB 92 IADVDKPFGL-PANLHWGLRSVASDGLDSMGHPRAIMGLTYCHG-----GNNQ 144
QY 502 VFTFTREDIRPDPOHTKKPCFDALSHSPVTLVYCHSMKNQWTKRK-DKTLVHVS 560
DB 145 LVRLNAGQLGVEER-----CVEA--DROGILKAVCRGLGVDPGPMQYNEHTGILMRVH 196
QY 561 GSGMDCSESDHRIFMNTCNPSLSLQWLFEE 590
DB 197 KCKMALHPATQQLSLGHCDVNDSYQWMPK 226

RESULT 36

US-09-270-767-61773
; Sequence 61773; Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 61773
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-61773

Query Match 7.0%; Score 228.5; DB 4; Length 290;
Best Local Similarity 28.6%; Pred. No. 1.1e-15;
Matches 82; Conservative 31; Mismatches 95; Indels 79; Gaps 14;

QY 345 EOYEISFKVMCGRMEDIPCSRGHYRKYVPYKVPAGVS--LARNLKRAEVMWDEYA 402
DB 1 ENWEMSRRIQCGGRVBISSCHGHVFRSSTPTFGCGMSEVLTDLAALATYMDW- 59
QY 403 EYIYORPEYRHLSAGD--VAVOKKRLSSLNCSKSPFWMTKIAWDLPKFYPPVEPPAA 459
DB 60 QYRIMUTSGLTGAKKAVTERVALRERLQCKPFSWYLENI-W--PEHFFPA--PDRF 114
QY 460 WGEIRNV-GTGLCA-----DTKGLGALGSLRLLE--GCVR 490
DB 115 FGIKIWDGETECQAYSKMKMLPGRLSRBWKRAFEEDSKAEELMALIDLRDKCLR 174
QY 491 -----GRGEAAMNMQVFTFTREDIRPDGP-----QHTK-----KFCFD 525
DB 175 PLKEDVRSLSAVTVDDCTSHAGSMQMFVITPRGQIMTNDVCLTRYRQOKLGVIKRLKN 234
QY 526 AISHTSVTLVYCHSMKNQWTKRKDKTLVHVS GSGMDCSESDHRI 572
DB 235 RNATTSNVMLAOCAS--DSSQLMTYD-----MDTQOQISHR 267

RESULT 37

US-09-270-767-61989
; Sequence 61989; Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 61989
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-61989

Query Match 6.7%; Score 218.5; DB 4; Length 72;
Best Local Similarity 56.2%; Pred. No. 1.3e-15;
Matches 41; Conservative 11; Mismatches 20; Indels 1; Gaps 1;

QY 218 IRTMLGASVATGDTTFLDSHCEANVNLPPILDRIRNRKTIYCPMIDVIDHDDFRYE 277
DB 1 IRARLGAHVSGEVITFLDACECTEGWLEPLARIVQNRRTVCPIIDVISDEFEYI 60
QY 278 TQAGDAMGAFDW 290
DB 61 T-ASDSTWGFNW 72

RESULT 38

US-09-270-767-33350
; Sequence 33350; Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 33350
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-33350

Query Match 5.1%; Score 168; DB 4; Length 73;
Best Local Similarity 47.5%; Pred. No. 3.6e-10;
Matches 29; Conservative 11; Mismatches 19; Indels 2; Gaps 1;

QY 352 KYMCCGGRMEDIPCSRGHYRKYVPYKVPAGVS--LARNLKRAEVMWDEYAEYIYORR 409
DB 12 QVMCCGVLEIAPCSKHVGRKSTPTFTFGGTTETIVNNARLVAVWLDWKEFYTSFY 71
QY 410 P 410
DB 72 P 72
RESULT 39
US-09-270-767-48567
; Sequence 48567; Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517

SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 48567
LENGTH: 73
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-09-270-767-48567

Query Match
Best Local Similarity 47.5%; Score 168; DB 4; Length 73;
Matches 29; Conservative 11; Mismatches 19; Indels 2; Gaps 1;

QY 352 KVMCGRMEDIPCSRVGHITRYKVPAGVS--LARNLKVAEVMDEVAEYIY 409
DB 12 QVMCGVLEIAPCSRVGHIFRDKSPYTFPGGTETVNNHNAVLVEWLDWKEFYSPY 71

QY 410 P 410
DB 72 P 72

RESULT 40
US-09-270-767-33441
Sequence 33441, Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:

APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 33441
LENGTH: 101
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-09-270-767-33441

Query Match
Best Local Similarity 48.3%; Score 167; DB 4; Length 101;
Matches 29; Conservative 12; Mismatches 17; Indels 2; Gaps 1;

QY 349 ISFKVMCGRMEDIPCSRVGHITRYKVPAGVS--LARNLKVAEVMDEVAEYIY 406
DB 26 INFQIWCGGILEITICSHVGHVFRDKSPYTFPGVAKIVLHNAARVAEVMDEWDFPY 85

RESULT 41
US-09-270-767-48658
Sequence 48658, Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:

APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 48658
LENGTH: 101
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-09-270-767-48658

Query Match
Best Local Similarity 5.1%; Score 167; DB 4; Length 101;
Matches 29; Conservative 12; Mismatches 17; Indels 2; Gaps 1;

QY 349 ISFKVMCGRMEDIPCSRVGHITRYKVPAGVS--LARNLKVAEVMDEVAEYIY 406
DB 26 INFQIWCGGILEITICSHVGHVFRDKSPYTFPGVAKIVLHNAARVAEVMDEWDFPY 85

RESULT 42
US-09-134-000C-5947
Sequence 5947, Application US/09134000C
Patent No. 6617156
GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
FILE REFERENCE: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
CURRENT APPLICATION NUMBER: US/09/134,000C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/055,778
PRIOR FILING DATE: 1997-08-15
NUMBER OF SEQ ID NOS: 6812
SOFTWARE: Patentin version 3.1
SEQ ID NO 5947
LENGTH: 424
TYPE: PRT
ORGANISM: Enterococcus faecalis
US-09-134-000C-5947

Query Match
Best Local Similarity 4.1%; Score 133; DB 4; Length 424;
Matches 59; Conservative 35; Mismatches 96; Indels 54; Gaps 7;

QY 142 ETLPTSTIIPPHNGWSLRTVSHVLSRSPPELVAELVDPDSDREHL-----KK 194
DB 51 EDAPFVSIVPAHNEG-IYIVATVEALNFNPQDRYEITIVINDSSDSASALLANVQCK 109

QY 195 PLEDYMALEPFSVRLITRTKREGILTRMKGASVATGDTVITFLDSCCANVNLPLLDR 254
DB 110 AVNRMLQITNDINIGKSGKSNALN---IGFKQAKGDVIAIYDADNTPPNALRYLVGEL 166

QY 255 -----ANRRTIYCPMIDVIDHDHDFRFEYQAGAMGAPFMEVYKRIRIP 300
DB 167 IASBEYGAIVGKFRTRNRASLLTRFINETLAFQWMAQGRF-----OLFUKCTIP 218

QY 301 PELQKADPEDPESPVMAGLPAVDKRFMELEGYDPGLEIMGBOYBISFKVMCGRM 360
DB 219 -----GTNFIYRSIIEALIGCHDDKAL---AEDTEISFRIYMGYKI 257

QY 361 EDIP 364
DB 258 KEOP 261

RESULT 43
US-09-270-767-33601
Sequence 33601, Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:

APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 33601
LENGTH: 88
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-09-270-767-33601

Query Match
Best Local Similarity 3.9%; Score 129; DB 4; Length 88;
Matches 26; Conservative 13; Mismatches 34; Indels 2; Gaps 2;

QY 531 SPVTLYDCHSMKGNOLWKY-RDKXTLYHPVSG-SCMDCSESDHRIFMNTCNSSLTQWL 588
DB 9 APVWLMDCHSGGANGYWIYDYRHKQLKGTGGRCLLELPFSQEVVANKCDYDNRPOQWN 68

Db 181 PEEVVVADDS--KENLITVQKEOKL--DIKYRQDGYQLCAVNLGRTAKYFVSI 238
QY 236 LDBHCANVWLPPLDRIARNKTIKVIDID-----HDDPEYEQADAR 285
Db 239 LDDCMAPOQLWVSYTELEDDIVLIGPKVYDTHNTAEOFLNDPILIESLPETATN 298
QY 286 G-----AFDMEYKRIPIPELOKADSPDESP--VMAGLPAVDRKWEIG 333
Db 299 NNPSTSKNISLDMRLEH-----FKTDLRLCDSPFRFVAGNVAFSKEMLNKVG 350
QY 334 GNDPGLIENGGEYEIFKVM--CGRMED-----IP 364
Db 351 WPEEBNHWGGEVEGEGYRLFAGCFRVIDGMAIHOEPKENEETEREAGKSITLKIY 410
QY 365 CSRVGHIYRYVP-----YKVPAGVSLANLKRVAEVMWDEY--AEYI 405
Db 411 KEKVPIYRKLPIEDSHIRIP-----LVSIYPAVNCANYI 448

RESULT 48
US-09-437-277-1
Sequence 1, Application US/09437277
Patent No. 6444447
GENERAL INFORMATION:
APPLICANT: THE BOARD OF REGENTS OF THE UNIVERSITY OF OKLAHOMA
TITLE OF INVENTION: POLYMER GRAFTING BY POLYSACCHARIDE SYNTHASES
FILE REFERENCE: 5820.551
CURRENT APPLICATION NUMBER: US/09/437, 277
CURRENT FILING DATE: 1999-11-10
NUMBER OF SEQ ID NOS: 6
SOFTWARE: WordPerfect 8.0 (saved in ASCII format)
SEQ ID NO 1
LENGTH: 702
TYPE: PRF
ORGANISM: Pasteurella multocida
US-09-437-277-1

Query Match 3.8%; Score 124; DB 4; Length 702;
Best Local Similarity 22.1%; Pred. No. 0.0089;
Matches 83; Conservative 49; Mismatches 156; Indels 88; Gaps 19;
QY 61 KTFELDGOK--LKD-WH--DKEAIRDAQVNGEGRPMTDAERVDQVRENGFN 114
Db 84 KKLVLSDSEKNTLKNKKLLEKKSSENAEVAVALVPKPPKDLVAPLDHV---NDFT 140
QY 115 IYVSDKISLNRSLPIRHPNCKRYLETLPNTSIIIPFNEGWSLRTVHSLNRSP 174
Db 141 WYKRRKRLGK-PEHOH-----VGLSIIVTFNR--PALSTTLACLNVOKT 185
QY 175 ELVAEIVLVDDFDRHLKKPLEDYMALPSPVRLRTKREG--LRTKMLGASVATGV 232
Db 186 HPEFEIYVDDGS-QEDLSIIIOYENKL-DIYVR-QKDNQQAARWGLRLAKYDF 242
QY 233 IFFLDHCHCANVWLPPLDRIARNK-TIVCP--MIDV--IDHDF-----RY 276
Db 243 IGLDDCDMAPNPLWVHVSVALEDDDLTIGRKXIDTQHIDPKOFLNNAJLSLEPVEY 302
QY 277 ETQAGAMG---AFDMEYKRIPIPELOKADSPDESP--VMAGLPAVDRKWEI 330
Db 303 KTNNSVAAGKEGTVSLDMRL-----BOFEKTENLRUSDSPFRFAGNVAFARKWN 354
QY 331 ELGGYDPGLIENGGEYEIFKVMCGRME-----361
Db 355 KSGPFDEEBNHWGGEVEGEGYRLFYGSFKTIDGIMAHQEPKENEETEREAGKNITL 414
QY 362 DIPCSRVGHIYRYVP 377
Db 415 DIMREKVPIYRKLPI 430

RESULT 49
US-09-134-000C-5087

Sequence 5087, Application US/09134000C
Patent No. 6617156
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 032796-032
CURRENT APPLICATION NUMBER: US/09/134,000C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/055,778
PRIOR FILING DATE: 1997-08-15
NUMBER OF SEQ ID NOS: 6812
SOFTWARE: PatentIn version 3.1
SEQ ID NO 5087
LENGTH: 534
TYPE: PRF
ORGANISM: Enterococcus faecalis
US-09-134-000C-5087

Query Match 3.8%; Score 123.5; DB 4; Length 534;
Best Local Similarity 22.7%; Pred. No. 0.0064;
Matches 56; Conservative 43; Mismatches 93; Indels 55; Gaps 11;
QY 145 PNTSIIIPPHN--EGWSLRTVHSLNRSPPELVAEIYVDDFDRHLKKPLEDYMAL 202
Db 193 PXSISAMPYVYNEBKRL--CIDSILNOVYTNW--ELCMADASTDPVVKILTEYOOL 248
QY 203 PPSVRLRTKREGIIRTMGLASVATGVITFLDHCANVWLPPLDRIARNKTIY 262
Db 249 DERIRVFERQNGHISEATNSALATGFEVLLDDELAINAYEVVKLVNEN----- 303
QY 263 CEMIDVIDHDDFRYETQADAMRGAFDMEYKRIPIPELOKADPS-DPESPVWAGG- 320
Db 304 -PELDIYSDDEKID-----NDG-----NRSDPAFRDMSPLDILGT 339
QY 321 ----LFAVDRKFMELGIDPGLIENGGEYEIFKVMCGRMEDIPCSRVGHI----- 371
Db 340 NYISHLGYVRRSILBIEGFRKGYE--GSQDYDLVLRF-----TEKTKERIKHPIKVL 391
QY 372 -YRKVYP 377
Db 392 YTRWLP 398

RESULT 50
US-09-107-532A-6889
Sequence 6889, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESSES:
ADDRESSER: GENOME THERAPEUTICS CORPORATION
CITY: Maitland
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997

```

ATTORNEY/AGENT INFORMATION:
NAME: Atinelli, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 6889:
SEQUENCE CHARACTERISTICS:
LENGTH: 721 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEetical: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1..721
SEQUENCE DESCRIPTION: SEQ ID NO: 6889:
IS-09-107-352A-6889

```

[illegible]

Search completed: November 22, 2004, 15:39:50
Job time : 49 secs

Inis Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2004, 14:04:27 / Search time 147 Seconds

(without alignments)
1467.854 Million cell updates/sec

Title: US-10-001-851-2

Perfect score: 3278

Sequence: 1 MRKEKRLQAVLVALV.....TQOWLEHTNSTVLEKENRN 603

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1589859 seqs, 357834939 residues

Total number of hits satisfying chosen parameters: 1589859

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

Published Applications AA:
1: /cgn2_6/prodata/1/pubppa/US07_PUBCOMB.pep:*
2: /cgn2_6/prodata/1/pubppa/PCT_NEW_PUB.pep:*
3: /cgn2_6/prodata/1/pubppa/US06_NEW_PUB.pep:*
4: /cgn2_6/prodata/1/pubppa/US06_PUBCOMB.pep:*
5: /cgn2_6/prodata/1/pubppa/US07_NEW_PUB.pep:*
6: /cgn2_6/prodata/1/pubppa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/prodata/1/pubppa/US08_NEW_PUB.pep:*
8: /cgn2_6/prodata/1/pubppa/US08_PUBCOMB.pep:*
9: /cgn2_6/prodata/1/pubppa/US09A_PUBCOMB.pep:*
10: /cgn2_6/prodata/1/pubppa/US09B_PUBCOMB.pep:*
11: /cgn2_6/prodata/1/pubppa/US09C_PUBCOMB.pep:*
12: /cgn2_6/prodata/1/pubppa/US09_NEW_PUB.pep:*
13: /cgn2_6/prodata/1/pubppa/US10A_PUBCOMB.pep:*
14: /cgn2_6/prodata/1/pubppa/US10B_PUBCOMB.pep:*
15: /cgn2_6/prodata/1/pubppa/US10C_PUBCOMB.pep:*
16: /cgn2_6/prodata/1/pubppa/US10D_PUBCOMB.pep:*
17: /cgn2_6/prodata/1/pubppa/US10_NEW_PUB.pep:*
18: /cgn2_6/prodata/1/pubppa/US11_NEW_PUB.pep:*
19: /cgn2_6/prodata/1/pubppa/US60_NEW_PUB.pep:*
20: /cgn2_6/prodata/1/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3278	100.0	603	US-09-795-926-43	Sequence 43, Appl
2	3278	100.0	603	US-10-001-851-2	Sequence 2, Appl
3	3278	100.0	603	US-10-364-774-43	Sequence 43, Appl
4	3242.5	98.9	631	US-09-795-926-41	Sequence 41, Appl
5	3242.5	98.9	631	US-10-364-774-41	Sequence 41, Appl
6	2771	84.5	506	US-09-795-926-31	Sequence 31, Appl
7	2771	84.5	506	US-10-364-774-31	Sequence 31, Appl
8	2746.5	83.8	535	US-09-795-926-29	Sequence 29, Appl
9	2746.5	83.8	535	US-10-364-774-29	Sequence 29, Appl
10	1877	57.3	366	US-09-795-926-39	Sequence 39, Appl
11	1877	57.3	366	US-10-364-774-39	Sequence 39, Appl
12	1392	42.5	276	US-10-292-896-3	Sequence 3, Appl
13	1370	41.8	269	US-09-795-926-27	Sequence 27, Appl

14	1370	41.8	269	US-10-364-774-27	Sequence 27, Appl
15	1311	40.0	321	US-09-795-926-35	Sequence 35, Appl
16	1311	40.0	321	US-10-364-774-35	Sequence 35, Appl
17	1125	34.3	561	US-09-925-301-1006	Sequence 1006, Ap
18	1117	34.1	559	US-10-001-851-24	Sequence 24, Appl
19	1116	34.0	559	US-10-001-851-21	Sequence 21, Appl
20	1116	34.0	559	US-10-205-219-36	Sequence 36, Appl
21	1116	34.0	559	US-10-205-219-76	Sequence 76, Appl
22	1115	34.0	559	US-10-001-851-22	Sequence 22, Appl
23	1115	34.0	559	US-10-001-851-23	Sequence 23, Appl
24	1113	34.0	559	US-10-001-851-20	Sequence 20, Appl
25	1106	33.7	556	US-10-292-896-62	Sequence 62, Appl
26	1079	32.9	626	US-10-001-851-27	Sequence 27, Appl
27	1050	32.0	187	US-10-292-896-118	Sequence 118, App
28	1033	31.5	578	US-10-074-527-8	Sequence 8, Appl
29	1029	31.4	657	US-10-341-434-32	Sequence 32, Appl
30	1013	30.9	940	US-10-821-273-86	Sequence 86, Appl
31	999.5	30.5	558	US-10-292-896-66	Sequence 66, Appl
32	999.5	30.5	558	US-10-433-256-6	Sequence 6, Appl
33	991.5	30.2	581	US-10-074-527-2	Sequence 2, Appl
34	991.5	30.2	581	US-10-085-198-12	Sequence 122, App
35	991.5	30.2	581	US-10-433-256-4	Sequence 4, Appl
36	990	30.2	581	US-10-789-241-44	Sequence 44, Appl
37	990	30.2	240	US-09-795-926-33	Sequence 33, Appl
38	990	30.2	240	US-10-364-774-33	Sequence 33, Appl
39	988.5	30.2	612	US-10-001-851-25	Sequence 25, Appl
40	988	30.1	209	US-09-795-926-37	Sequence 37, Appl
41	988	30.1	209	US-10-364-774-37	Sequence 37, Appl
42	971.5	29.6	581	US-10-292-896-58	Sequence 58, Appl
43	971	29.6	552	US-09-815-028-2	Sequence 2, Appl
44	971	29.6	552	US-10-028-072-196	Sequence 196, App
45	971	29.6	552	US-10-140-808-196	Sequence 196, App

ALIGNMENTS

RESULT 1
US-09-795-926-43
Sequence 43, Application US/09795926
Patent No. US2002098486A1
GENERAL INFORMATION:
APPLICANT: Donoho, Gregory
APPLICANT: Hilbun, Brian
APPLICANT: Turner, C. Alexander Jr.
APPLICANT: Friedrich, Glenn
APPLICANT: Abulin, Alejandro
APPLICANT: Zambrowicz, Brian
APPLICANT: Sands, Arthur T.
APPLICANT: Walke, D. Wade
APPLICANT: Wilgenowski, Nathaniel L.
APPLICANT: Hu, Yi
APPLICANT: Kieke, James Alvin
APPLICANT: Poter, David George
TITLE OF INVENTION: NOVEL HUMAN TRANSEPERASE PROTEINS AND
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME
FILE REFERENCE: LEX-0144-USA
CURRENT APPLICATION NUMBER: US/09/795,926
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 60/185,920
PRIOR FILING DATE: 2000-02-29
PRIOR APPLICATION NUMBER: US 60/186,558
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: US 60/191,849
NUMBER OF SEQ ID NOS: 47
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 43
LENGTH: 603
TYPE: PRT
ORGANISM: homo sapiens
US-09-795-926-43

Query Match 100.0%; Score 3278; DB 9; Length 603;
Best Local Similarity 100.0%; Pred. No. 5e-310;
Matches 603; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKREKRLLOAVLVAALVLLPNVGLMALYRERODGTTPGSGAAVAPAAAGGSHSRK 60
DB 1 MRKREKRLLOAVLVAALVLLPNVGLMALYRERODGTTPGSGAAVAPAAAGGSHSRK 60
QY 61 KTFPLDGGKLDKMDHKEAIRDAQRVNGEGRPYPMTDARVDAQYRENGFNIVSDK 120
DB 61 KTFPLDGGKLDKMDHKEAIRDAQRVNGEGRPYPMTDARVDAQYRENGFNIVSDK 120
QY 121 ISLNSRLPIRHPNCSKRYLETLPNTSIIIPHNAGWSLLRTVSHVNSRSPPELVAI 180
DB 121 ISLNSRLPIRHPNCSKRYLETLPNTSIIIPHNAGWSLLRTVSHVNSRSPPELVAI 180
QY 181 VLVDPDSDBEHLKKPLEDMALPPSVRLIRTKRREGILRTMLGASVATGDVITFLDSHC 240
DB 181 VLVDPDSDBEHLKKPLEDMALPPSVRLIRTKRREGILRTMLGASVATGDVITFLDSHC 240
QY 241 EAVNVMPLPLDLRIANRKTIVCPMIDVIDHDDFRYETQAGDAMGAPDMEMYKRIPI 300
DB 241 EAVNVMPLPLDLRIANRKTIVCPMIDVIDHDDFRYETQAGDAMGAPDMEMYKRIPI 300
QY 301 PELQKADPSDPFESPVMAGGLPAVDRKMFELGSDYDGLGELWGEQYEISFKVMCGGM 360
DB 301 PELQKADPSDPFESPVMAGGLPAVDRKMFELGSDYDGLGELWGEQYEISFKVMCGGM 360
QY 361 EDIPCSRVGHIYRKVPYVPAGVSLARNLKRYAEVWDEYAEYIYQRPBRYHLSAGDV 420
DB 361 EDIPCSRVGHIYRKVPYVPAGVSLARNLKRYAEVWDEYAEYIYQRPBRYHLSAGDV 420
QY 421 AVQKRLSSLNCKSFPMFTKIAMDLPKFYPPVEPAAAMGELRVNGTGLCADTKHGALG 480
DB 421 AVQKRLSSLNCKSFPMFTKIAMDLPKFYPPVEPAAAMGELRVNGTGLCADTKHGALG 480
QY 481 SPLRLGECYRGGEAAMNNMQUFTFTWRREDIRPGDPQHTKKCFDAISHTSPVTLVDCHS 540
DB 481 SPLRLGECYRGGEAAMNNMQUFTFTWRREDIRPGDPQHTKKCFDAISHTSPVTLVDCHS 540
QY 541 MGNQMLMKRKDKTLYHPVSGSCMDCSBDHRI FNNTCNPSSLTQOMLFEHTNSTVLEKF 600
DB 541 MGNQMLMKRKDKTLYHPVSGSCMDCSBDHRI FNNTCNPSSLTQOMLFEHTNSTVLEKF 600
QY 601 NRN 603
DB 601 NRN 603

RESULT 2
US-10-001-851-2
Sequence 2, Application US/10001851
Publication No. US20020115628A1
GENERAL INFORMATION:
APPLICANT: MEYERS, Rachel A.
APPLICANT: WILLIAMSON, Mark
TITLE OF INVENTION: 47169 and 33935, No. US20020115628A1el Human Glycosyl Transferase
TITLE OF INVENTION: Uses Thereof
FILE REFERENCE: 10147-56U1
CURRENT APPLICATION NUMBER: US/10/001,851
CURRENT FILING DATE: 2001-11-20
PRIOR APPLICATION NUMBER: US 60/249,939
PRIOR FILING DATE: 2000-11-20
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 603
TYPE: PRT
ORGANISM: Homo sapiens
US-10-001-851-2

Query Match 100.0%; Score 3278; DB 13; Length 603;
Best Local Similarity 100.0%; Pred. No. 5e-310;

Matches 603; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKREKRLLOAVLVAALVLLPNVGLMALYRERODGTTPGSGAAVAPAAAGGSHSRK 60
DB 1 MRKREKRLLOAVLVAALVLLPNVGLMALYRERODGTTPGSGAAVAPAAAGGSHSRK 60
QY 61 KTFPLDGGKLDKMDHKEAIRDAQRVNGEGRPYPMTDARVDAQYRENGFNIVSDK 120
DB 61 KTFPLDGGKLDKMDHKEAIRDAQRVNGEGRPYPMTDARVDAQYRENGFNIVSDK 120
QY 121 ISLNSRLPIRHPNCSKRYLETLPNTSIIIPHNAGWSLLRTVSHVNSRSPPELVAI 180
DB 121 ISLNSRLPIRHPNCSKRYLETLPNTSIIIPHNAGWSLLRTVSHVNSRSPPELVAI 180
QY 181 VLVDPDSDBEHLKKPLEDMALPPSVRLIRTKRREGILRTMLGASVATGDVITFLDSHC 240
DB 181 VLVDPDSDBEHLKKPLEDMALPPSVRLIRTKRREGILRTMLGASVATGDVITFLDSHC 240
QY 241 EAVNVMPLPLDLRIANRKTIVCPMIDVIDHDDFRYETQAGDAMGAPDMEMYKRIPI 300
DB 241 EAVNVMPLPLDLRIANRKTIVCPMIDVIDHDDFRYETQAGDAMGAPDMEMYKRIPI 300
QY 301 PELQKADPSDPFESPVMAGGLPAVDRKMFELGSDYDGLGELWGEQYEISFKVMCGGM 360
DB 301 PELQKADPSDPFESPVMAGGLPAVDRKMFELGSDYDGLGELWGEQYEISFKVMCGGM 360
QY 361 EDIPCSRVGHIYRKVPYVPAGVSLARNLKRYAEVWDEYAEYIYQRPBRYHLSAGDV 420
DB 361 EDIPCSRVGHIYRKVPYVPAGVSLARNLKRYAEVWDEYAEYIYQRPBRYHLSAGDV 420
QY 421 AVQKRLSSLNCKSFPMFTKIAMDLPKFYPPVEPAAAMGELRVNGTGLCADTKHGALG 480
DB 421 AVQKRLSSLNCKSFPMFTKIAMDLPKFYPPVEPAAAMGELRVNGTGLCADTKHGALG 480
QY 481 SPLRLGECYRGGEAAMNNMQUFTFTWRREDIRPGDPQHTKKCFDAISHTSPVTLVDCHS 540
DB 481 SPLRLGECYRGGEAAMNNMQUFTFTWRREDIRPGDPQHTKKCFDAISHTSPVTLVDCHS 540
QY 541 MGNQMLMKRKDKTLYHPVSGSCMDCSBDHRI FNNTCNPSSLTQOMLFEHTNSTVLEKF 600
DB 541 MGNQMLMKRKDKTLYHPVSGSCMDCSBDHRI FNNTCNPSSLTQOMLFEHTNSTVLEKF 600
QY 601 NRN 603
DB 601 NRN 603

RESULT 3
US-10-364-774-43
Sequence 43, Application US/10364774
Publication No. US20030144497A1
GENERAL INFORMATION:
APPLICANT: Donoho, Gregory
APPLICANT: Hilbun, Erin
APPLICANT: Turner, C. Alexander Jr.
APPLICANT: Friedrich, Glenn
APPLICANT: Abuh, Alejandro
APPLICANT: Zambrowicz, Brian
APPLICANT: Sander, Arthur T.
APPLICANT: Walke, D. Wade
APPLICANT: Wilganowski, Nathaniel L.
APPLICANT: Hu, Yi
APPLICANT: Kleke, James Alvin
APPLICANT: Potter, David George
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME
FILE REFERENCE: LEX-0144-USA
CURRENT APPLICATION NUMBER: US/10/364,774
CURRENT FILING DATE: 2003-02-11
PRIOR APPLICATION NUMBER: US/09/795,926
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 60/185,920
PRIOR FILING DATE: 2000-02-29

; PRIOR APPLICATION NUMBER: US 60/186,558
 ; PRIOR FILING DATE: 2000-03-02
 ; PRIOR APPLICATION NUMBER: US 60/191,849
 ; PRIOR FILING DATE: 2000-03-24
 ; NUMBER OF SEQ ID NOS: 47
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 43
 ; LENGTH: 603
 ; TYPE: PRF
 ; ORGANISM: homo sapiens
 ; US-10-364-774-43

Query Match 100.0%; Score 3278; DB 14; Length 603;
 Best Local Similarity 100.0%; Pred. No. 5e-310;
 Matches 603; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRRERKRLQAVALVLAALVLLPNVGMALYRERQDPTPGSGAAYAPAAAGGSHSRK 60
 DB 1 MRRERKRLQAVALVLAALVLLPNVGMALYRERQDPTPGSGAAYAPAAAGGSHSRK 60
 QY 61 KTFPLDGGQKLKDMHDEKAI RDAQVNGEGORPYPMTDARVDQAYRENGFNITYSDK 120
 DB 61 KTFPLDGGQKLKDMHDEKAI RDAQVNGEGORPYPMTDARVDQAYRENGFNITYSDK 120
 QY 121 ISLNRSLPDIRHPCNSKRYLETLPNTSIIIPFHNGMSSLRTVHSVLRSPPELVAEI 180
 DB 121 ISLNRSLPDIRHPCNSKRYLETLPNTSIIIPFHNGMSSLRTVHSVLRSPPELVAEI 180
 QY 181 VLVDPSDBREHLKKPLEDYMALFPVSRIILRTKKREGILRTMLGASVATGDTITFLDSHC 240
 DB 181 VLVDPSDBREHLKKPLEDYMALFPVSRIILRTKKREGILRTMLGASVATGDTITFLDSHC 240
 QY 241 EAVVNMPLPLLDRIARNRKTIYCPMTDIVDHDFFRYETQAGDAMRGAFPMEMYKRIPI 300
 DB 241 EAVVNMPLPLLDRIARNRKTIYCPMTDIVDHDFFRYETQAGDAMRGAFPMEMYKRIPI 300
 QY 301 PELQKADPSDPFSPVMAAGLFAVDRKFMELGSDYDGLIINGGEOYEISFKYMMCGGRM 360
 DB 301 PELQKADPSDPFSPVMAAGLFAVDRKFMELGSDYDGLIINGGEOYEISFKYMMCGGRM 360
 QY 361 EDIPCSRVGHIYRKYVPYKVPAGVSLARNLKRVAEVMWDEYAEYIYQREBYEHLISAGDV 420
 DB 361 EDIPCSRVGHIYRKYVPYKVPAGVSLARNLKRVAEVMWDEYAEYIYQREBYEHLISAGDV 420
 QY 421 AVQKRLRSSLNCSSFKFMFTKIAMDLKFPYVPPPAANGERTINVTGICADTKGALG 480
 DB 421 AVQKRLRSSLNCSSFKFMFTKIAMDLKFPYVPPPAANGERTINVTGICADTKGALG 480
 QY 481 SPLRLEGCVARGGEAANNNOVFTWRREDIRPGDPQHTKKPCPDASHTSPTLYDCHS 540
 DB 481 SPLRLEGCVARGGEAANNNOVFTWRREDIRPGDPQHTKKPCPDASHTSPTLYDCHS 540
 QY 541 MKGNOLMKYRKDKTYLHPVSGSCMDSCSDHRIFMNTCNPSSTLOQWLFHTNSTVLEKF 600
 DB 541 MKGNOLMKYRKDKTYLHPVSGSCMDSCSDHRIFMNTCNPSSTLOQWLFHTNSTVLEKF 600
 QY 601 NRN 603
 DB 601 NRN 603
 RESULT 4
 US-09-795-926-41
 ; Sequence 41, Application US/09795926
 ; Patent No. US20020098486A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Donoho, Gregory
 ; APPLICANT: Hilbun, Brian
 ; APPLICANT: Turner, C. Alexander Jr.
 ; APPLICANT: Friedlich, Glenn
 ; APPLICANT: Aduin, Alejandro
 ; APPLICANT: Zambrowicz, Brian
 ; APPLICANT: Sands, Arthur T.

; APPLICANT: Walke, D. Wade
 ; APPLICANT: Wilgenowski, Nathaniel L.
 ; APPLICANT: Hu, Yi
 ; APPLICANT: Kleke, James Alvin
 ; APPLICANT: Potter, David George
 ; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
 ; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME
 ; FILE REFERENCE: LEX-0144-USA
 ; CURRENT APPLICATION NUMBER: US/09/795,926
 ; CURRENT FILING DATE: 2001-02-28
 ; PRIOR APPLICATION NUMBER: US 60/185,920
 ; PRIOR FILING DATE: 2000-02-29
 ; PRIOR APPLICATION NUMBER: US 60/186,558
 ; PRIOR FILING DATE: 2000-03-02
 ; PRIOR APPLICATION NUMBER: US 60/191,849
 ; PRIOR FILING DATE: 2000-03-24
 ; NUMBER OF SEQ ID NOS: 47
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 41
 ; LENGTH: 631
 ; TYPE: PRF
 ; ORGANISM: homo sapiens
 ; US-09-795-926-41

Query Match 98.9%; Score 3242.5; DB 9; Length 631;
 Best Local Similarity 95.4%; Pred. No. 1.6e-306;
 Matches 601; Conservative 0; Mismatches 0; Indels 29; Gaps 1;

QY 1 MRRERKRLQAVALVLAALVLLPNVGMALYRERQDPTPGSGAAYAPAAAGGSHSRK 60
 DB 1 MRRERKRLQAVALVLAALVLLPNVGMALYRERQDPTPGSGAAYAPAAAGGSHSRK 60
 QY 61 KTFPLDGGQKLKDMHDEKAI RDAQVNGEGORPYPMTDARVDQAYRENGFNITYSDK 120
 DB 61 KTFPLDGGQKLKDMHDEKAI RDAQVNGEGORPYPMTDARVDQAYRENGFNITYSDK 120
 QY 121 ISLNRSLPDIRHPCNSKRYLETLPNTSIIIPFHNGMSSLRTVHSVLRSPPELVAEI 180
 DB 121 ISLNRSLPDIRHPCNSKRYLETLPNTSIIIPFHNGMSSLRTVHSVLRSPPELVAEI 180
 QY 181 VLVDPSDBREHLKKPLEDYMALFPVSRIILRTKKREGILRTMLGASVATGDTITFLDSHC 240
 DB 181 VLVDPSDBREHLKKPLEDYMALFPVSRIILRTKKREGILRTMLGASVATGDTITFLDSHC 240
 QY 241 EAVVNMPLPLLDRIARNRKTIYCPMTDIVDHDFFRYETQAGDAMRGAFPMEMYKRIPI 300
 DB 241 EAVVNMPLPLLDRIARNRKTIYCPMTDIVDHDFFRYETQAGDAMRGAFPMEMYKRIPI 300
 QY 301 PELQKADPSDPFSPVMAAGLFAVDRKFMELGSDYDGLIINGGEOYEISFKYMMCGGRM 360
 DB 301 PELQKADPSDPFSPVMAAGLFAVDRKFMELGSDYDGLIINGGEOYEISFKYMMCGGRM 360
 QY 361 VSNMSPQAVFLPRAPMMLALQVMMCGGRMEDIPCSRVGHIYRKYVPYKVPAGVSLARNL 420
 DB 361 VSNMSPQAVFLPRAPMMLALQVMMCGGRMEDIPCSRVGHIYRKYVPYKVPAGVSLARNL 420
 QY 421 RVAEVMWDEYAEYIYQREBYEHLISAGDVAVOQKRLSSLNCSSFKFMFTKIAMDLKFPY 480
 DB 421 RVAEVMWDEYAEYIYQREBYEHLISAGDVAVOQKRLSSLNCSSFKFMFTKIAMDLKFPY 480
 QY 481 SPLRLEGCVARGGEAANNNOVFTWRREDIRPGDPQHTKKPCPDASHTSPTLYDCHS 540
 DB 481 SPLRLEGCVARGGEAANNNOVFTWRREDIRPGDPQHTKKPCPDASHTSPTLYDCHS 540
 QY 541 RPDGPQHTKKFCEDASHTSPTLYDCHSMKGNOLMKYRKDKTYLHPVSGSCMDSCSDH 571
 DB 541 RPDGPQHTKKFCEDASHTSPTLYDCHSMKGNOLMKYRKDKTYLHPVSGSCMDSCSDH 571
 QY 572 RIFMNTCNPSSTLOQWLFHTNSTVLEKFN 601
 DB 601 RIFMNTCNPSSTLOQWLFHTNSTVLEKFN 601

RESULT 5

US-10-364-774-41
 ; Sequence 41, Application US/10364774
 ; Publication No. US2003014497A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Donoho, Gregory
 ; APPLICANT: Hilbun, Erin
 ; APPLICANT: Turner, C. Alexander Jr.
 ; APPLICANT: Friedlich, Glenn
 ; APPLICANT: Abuin, Alejandro
 ; APPLICANT: Zambrowicz, Brian
 ; APPLICANT: Sands, Arthur T.
 ; APPLICANT: Walke, D. Wade
 ; APPLICANT: Wilganowski, Nathaniel L.
 ; APPLICANT: Hu, Yi
 ; APPLICANT: Kleke, James Alvin
 ; APPLICANT: Potter, David George
 ; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
 ; FILE REFERENCE: LEX-0144-USA
 ; CURRENT APPLICATION NUMBER: US/10/364,774
 ; PRIOR FILING DATE: 2003-02-11
 ; PRIOR APPLICATION NUMBER: US/09/795,926
 ; PRIOR FILING DATE: 2001-02-28
 ; PRIOR APPLICATION NUMBER: US 60/185,920
 ; PRIOR FILING DATE: 2000-02-29
 ; PRIOR APPLICATION NUMBER: US 60/186,558
 ; PRIOR FILING DATE: 2000-03-02
 ; PRIOR APPLICATION NUMBER: US 60/191,849
 ; PRIOR FILING DATE: 2000-03-24
 ; NUMBER OF SEQ ID NOS: 47
 ; SOFTWARE: FaastSeq for Windows Version 4.0
 ; SEQ ID NO 41
 ; LENGTH: 631
 ; TYPE: PRT
 ; ORGANISM: homo sapiens
 ; US-10-364-774-41

Query Match

Best Local Similarity 98.9%; Score 3242.5; DB 14; Length 631;
 Matches 60; Conservative 0; Mismatches 0; Indels 29; Gaps 1;

QY 1 MRREKRLLOVALVLAALVLLPVGIMALYRERQDPGTGGSGAAYAPAAAGGSHSRQK 60
 DB 1 MRREKRLLOVALVLAALVLLPVGIMALYRERQDPGTGGSGAAYAPAAAGGSHSRQK 60
 QY 61 KTFELDGGQKLKMDHDEAIRDAQRYNGEGRPYMTAEVDDAAYRNGENIYVSDK 120
 DB 61 KTFELDGGQKLKMDHDEAIRDAQRYNGEGRPYMTAEVDDAAYRNGENIYVSDK 120
 QY 121 ISLNRSIPDIRHNCNCKRYLETLPNTSIIIPHNESGSSILRTVSHVLSNRSPELVAET 180
 DB 121 ISLNRSIPDIRHNCNCKRYLETLPNTSIIIPHNESGSSILRTVSHVLSNRSPELVAET 180
 QY 181 VLVDPSDREHLKKPLEDYVALFPVSRIILTKKREGILRTMLGASVATGDTVTFLDSHC 240
 DB 181 VLVDPSDREHLKKPLEDYVALFPVSRIILTKKREGILRTMLGASVATGDTVTFLDSHC 240
 QY 241 EAVVMNLPPLLDIAIRNRKTIIVCPMTDIVDHPFRYETQAGDMRGAFFMEWYKXIP 300
 DB 241 EAVVMNLPPLLDIAIRNRKTIIVCPMTDIVDHPFRYETQAGDMRGAFFMEWYKXIP 300
 QY 301 PELQKADPSDPFESSPVMAAGLFAVDRKMFELGSGYDPGLEIMGGEQYEISFK----- 352
 DB 301 PELQKADPSDPFESSPVMAAGLFAVDRKMFELGSGYDPGLEIMGGEQYEISFK----- 352
 QY 353 -----VMMCGRMEDIFCSRVGHIYRKYPYKVPAGVSLARNLK 391
 DB 353 -----VMMCGRMEDIFCSRVGHIYRKYPYKVPAGVSLARNLK 391
 QY 392 RAVEVMNDEYAEYIYQRRPEYRHLASGDAVAVOKKLSSLNCKSPKFMNTKIANDLRFY 451
 DB 392 RAVEVMNDEYAEYIYQRRPEYRHLASGDAVAVOKKLSSLNCKSPKFMNTKIANDLRFY 451
 QY 421 RAVEVMNDEYAEYIYQRRPEYRHLASGDAVAVOKKLSSLNCKSPKFMNTKIANDLRFY 480
 DB 421 RAVEVMNDEYAEYIYQRRPEYRHLASGDAVAVOKKLSSLNCKSPKFMNTKIANDLRFY 480

RESULT 6

US-09-795-926-31
 ; Sequence 31, Application US/09795926
 ; Patent No. US20020098486A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Donoho, Gregory
 ; APPLICANT: Hilbun, Erin
 ; APPLICANT: Turner, C. Alexander Jr.
 ; APPLICANT: Friedlich, Glenn
 ; APPLICANT: Abuin, Alejandro
 ; APPLICANT: Zambrowicz, Brian
 ; APPLICANT: Sands, Arthur T.
 ; APPLICANT: Walke, D. Wade
 ; APPLICANT: Wilganowski, Nathaniel L.
 ; APPLICANT: Hu, Yi
 ; APPLICANT: Kleke, James Alvin
 ; APPLICANT: Potter, David George
 ; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
 ; FILE REFERENCE: LEX-0144-USA
 ; CURRENT APPLICATION NUMBER: US/09/795,926
 ; PRIOR FILING DATE: 2001-02-28
 ; PRIOR APPLICATION NUMBER: US 60/185,920
 ; PRIOR FILING DATE: 2000-02-29
 ; PRIOR APPLICATION NUMBER: US 60/186,558
 ; PRIOR FILING DATE: 2000-03-02
 ; PRIOR APPLICATION NUMBER: US 60/191,849
 ; PRIOR FILING DATE: 2000-03-24
 ; NUMBER OF SEQ ID NOS: 47
 ; SOFTWARE: FaastSeq for Windows Version 4.0
 ; SEQ ID NO 31
 ; LENGTH: 506
 ; TYPE: PRT
 ; ORGANISM: homo sapiens
 ; US-09-795-926-31

Query Match

Best Local Similarity 84.5%; Score 2771; DB 9; Length 506;
 Matches 506; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 MTDARVDDAYRNGENIYVSDKISLNRSIPDIRHNCNCKRYLETLPNTSIIIPHNESG 157
 DB 1 MTDARVDDAYRNGENIYVSDKISLNRSIPDIRHNCNCKRYLETLPNTSIIIPHNESG 157
 QY 158 WSSLRTVSHVLSNRSPELVAETVLVDPSDREHLKKPLEDYVALFPVSRIILTKKREG 217
 DB 158 WSSLRTVSHVLSNRSPELVAETVLVDPSDREHLKKPLEDYVALFPVSRIILTKKREG 217
 QY 218 IRTMLGASVATGDTVTFLDSHCANVVMNLPPLDRIANRRTIIVCPMTDIVDHPFRYE 277
 DB 218 IRTMLGASVATGDTVTFLDSHCANVVMNLPPLDRIANRRTIIVCPMTDIVDHPFRYE 277
 QY 278 TQAGDMRGAFFMEWYKXIPPELQKADPSDPFESSPVMAAGLFAVDRKMFELGSGYDP 337
 DB 278 TQAGDMRGAFFMEWYKXIPPELQKADPSDPFESSPVMAAGLFAVDRKMFELGSGYDP 337
 QY 338 GLEIMGGEQYEISFKVMCGRMEDIFCSRVGHIYRKYPYKVPAGVSLARNLKRYAEVW 397
 DB 338 GLEIMGGEQYEISFKVMCGRMEDIFCSRVGHIYRKYPYKVPAGVSLARNLKRYAEVW 397
 QY 421 GLEIMGGEQYEISFKVMCGRMEDIFCSRVGHIYRKYPYKVPAGVSLARNLKRYAEVW 480
 DB 421 GLEIMGGEQYEISFKVMCGRMEDIFCSRVGHIYRKYPYKVPAGVSLARNLKRYAEVW 480

QY 398 MDEAYIYORREPEYHLSAGDVAVOGKLRSSLNCKSPFKMFTKIAMDLPKFYPPVEPPA 457
 DB 301 MDEAYIYORREPEYHLSAGDVAVOGKLRSSLNCKSPFKMFTKIAMDLPKFYPPVEPPA 360
 QY 458 AAMGEIRNVGTGICADTKHGALGSPRLREGCVRGGEAAMNNQVFTFWREDIRGDDPQ 517
 DB 361 AAMGEIRNVGTGICADTKHGALGSPRLREGCVRGGEAAMNNQVFTFWREDIRGDDPQ 420
 QY 518 HTKKFCFDAISHTSPVTLVDCHSMKGNQMLKRYKDKTLVHPVSGSCMDCSESDHRIFFMNT 577
 DB 421 HTKKFCFDAISHTSPVTLVDCHSMKGNQMLKRYKDKTLVHPVSGSCMDCSESDHRIFFMNT 480
 QY 578 CNPSSLTQOQWLFPHNTSTVLEKFNRN 603
 DB 481 CNPSSLTQOQWLFPHNTSTVLEKFNRN 506

RESULT 7

US-10-364-774-31
 ; Sequence 31, Application US/10364774
 ; Publication No. US2003014497A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Donoho, Gregory
 ; APPLICANT: Hilbun, Erin
 ; APPLICANT: Turner, C. Alexander Jr.
 ; APPLICANT: Friedrich, Glenn
 ; APPLICANT: Abulin, Alejandro
 ; APPLICANT: Zambrowicz, Brian
 ; APPLICANT: Sands, Arthur T.
 ; APPLICANT: Walke, D. Wade
 ; APPLICANT: Wiganowski, Nathaniel L.
 ; APPLICANT: Hu, Yi
 ; APPLICANT: Kleke, James Alvin
 ; APPLICANT: Potter, David George
 ; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
 ; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME
 ; FILE REFERENCE: LEX-0144-USA
 ; CURRENT APPLICATION NUMBER: US/10/364,774
 ; CURRENT FILING DATE: 2003-02-11
 ; PRIOR APPLICATION NUMBER: US/09/795,926
 ; PRIOR FILING DATE: 2001-02-28
 ; PRIOR APPLICATION NUMBER: US 60/185,920
 ; PRIOR FILING DATE: 2000-02-29
 ; PRIOR APPLICATION NUMBER: US 60/186,558
 ; PRIOR FILING DATE: 2000-03-02
 ; PRIOR APPLICATION NUMBER: US 60/191,849
 ; PRIOR FILING DATE: 2000-03-24
 ; NUMBER OF SEQ ID NOS: 47
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 31
 ; LENGTH: 506
 ; TYPE: PRT
 ; ORGANISM: homo sapiens
 US-10-364-774-31

Query Match 84.5%; Score 2771; DB 14; Length 506;
 Best Local Similarity 100.0%; Pred. No. 1,le-260;
 Matches 506; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 MTAERVDQAYRENGENIYVSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPFNEG 157
 DB 1 MTAERVDQAYRENGENIYVSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPFNEG 60
 QY 158 WSSLIRTVHSLNRSPPELVAEIVLVDDFSDREHLKKPLEDYMALPPSVRIILTKKREG 217
 DB 61 WSSLIRTVHSLNRSPPELVAEIVLVDDFSDREHLKKPLEDYMALPPSVRIILTKKREG 120
 QY 218 IRTMLGASVATGVTITFLDSHCANVNLPLLDRIARNRKTIIVCPMIDVIDHDDFRYE 277
 DB 121 IRTMLGASVATGVTITFLDSHCANVNLPLLDRIARNRKTIIVCPMIDVIDHDDFRYE 180
 QY 278 TQAGDARGAFFDMWYTKRIPPELOKADSPDFESPWAGGLFAVDRKMFWEIGYDP 337

DB 181 TQAGDARGAFFDMWYTKRIPPELOKADSPDFESPWAGGLFAVDRKMFWEIGYDP 240
 QY 338 GLSITNGGEYEISFKWMCNGRMEDIPCSRVGHIYRKYPYKVPAGVSLARNLKRYAEW 397
 DB 241 GLSITNGGEYEISFKWMCNGRMEDIPCSRVGHIYRKYPYKVPAGVSLARNLKRYAEW 300
 QY 398 MDEAYIYORREPEYHLSAGDVAVOGKLRSSLNCKSPFKMFTKIAMDLPKFYPPVEPPA 457
 DB 301 MDEAYIYORREPEYHLSAGDVAVOGKLRSSLNCKSPFKMFTKIAMDLPKFYPPVEPPA 360
 QY 458 AAMGEIRNVGTGICADTKHGALGSPRLREGCVRGGEAAMNNQVFTFWREDIRGDDPQ 517
 DB 361 AAMGEIRNVGTGICADTKHGALGSPRLREGCVRGGEAAMNNQVFTFWREDIRGDDPQ 420
 QY 518 HTKKFCFDAISHTSPVTLVDCHSMKGNQMLKRYKDKTLVHPVSGSCMDCSESDHRIFFMNT 577
 DB 421 HTKKFCFDAISHTSPVTLVDCHSMKGNQMLKRYKDKTLVHPVSGSCMDCSESDHRIFFMNT 480
 QY 578 CNPSSLTQOQWLFPHNTSTVLEKFNRN 603
 DB 481 CNPSSLTQOQWLFPHNTSTVLEKFNRN 506

RESULT 8

US-09-795-926-29
 ; Sequence 29, Application US/09795926
 ; Patent No. US20020098486A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Donoho, Gregory
 ; APPLICANT: Hilbun, Erin
 ; APPLICANT: Turner, C. Alexander Jr.
 ; APPLICANT: Friedrich, Glenn
 ; APPLICANT: Abulin, Alejandro
 ; APPLICANT: Zambrowicz, Brian
 ; APPLICANT: Sands, Arthur T.
 ; APPLICANT: Walke, D. Wade
 ; APPLICANT: Wiganowski, Nathaniel L.
 ; APPLICANT: Hu, Yi
 ; APPLICANT: Kleke, James Alvin
 ; APPLICANT: Potter, David George
 ; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
 ; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME
 ; FILE REFERENCE: LEX-0144-USA
 ; CURRENT APPLICATION NUMBER: US/09/795,926
 ; CURRENT FILING DATE: 2001-02-28
 ; PRIOR APPLICATION NUMBER: US 60/185,920
 ; PRIOR FILING DATE: 2000-02-29
 ; PRIOR APPLICATION NUMBER: US 60/186,558
 ; PRIOR FILING DATE: 2000-03-02
 ; PRIOR APPLICATION NUMBER: US 60/191,849
 ; PRIOR FILING DATE: 2000-03-24
 ; NUMBER OF SEQ ID NOS: 47
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 29
 ; LENGTH: 535
 ; TYPE: PRT
 ; ORGANISM: homo sapiens
 US-09-795-926-29

Query Match 83.8%; Score 2746.5; DB 9; Length 535;
 Best Local Similarity 94.6%; Pred. No. 2,8e-258;
 Matches 506; Conservative 0; Mismatches 0; Indels 29; Gaps 1;

QY 98 MTAERVDQAYRENGENIYVSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPFNEG 157
 DB 1 MTAERVDQAYRENGENIYVSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPFNEG 60
 QY 158 WSSLIRTVHSLNRSPPELVAEIVLVDDFSDREHLKKPLEDYMALPPSVRIILTKKREG 217
 DB 61 WSSLIRTVHSLNRSPPELVAEIVLVDDFSDREHLKKPLEDYMALPPSVRIILTKKREG 120
 QY 218 IRTMLGASVATGVTITFLDSHCANVNLPLLDRIARNRKTIIVCPMIDVIDHDDFRYE 277

DB 121 IRTMLGASVATGDTVTFLDISHCEANVMWLPLLDRIARNRTIYCPMIDVIDHDDFRYE 180
DB 278 TOAGDMRGAFDWMYKRIPIPELOKADPSDPFSPVAGGLPAVDRKMFELGGYR 337
DB 181 TOAGDMRGAFDWMYKRIPIPELOKADPSDPFSPVAGGLPAVDRKMFELGGYR 240
QY 338 GLEIWGEQEYISFK-----VMMCGRMEDIPCSRY 368
DB 241 GLEIWGEQEYISFKGLHMLPRLVSNWPQAVFLPAPNMLAQVMMCGRMEDIPCSRY 300
QY 369 GHYRKYVPYKYPAGVSLARNLKRVAEVMWDEYAETIYORRPEYRHLASGDAVVOCKLRS 428
DB 301 GHYRKYVPYKYPAGVSLARNLKRVAEVMWDEYAETIYORRPEYRHLASGDAVVOCKLRS 360
QY 429 SLNCKSFKFMFTKIAMDLPKFPYPPVPPAAWGEIRNVGTGLCADTKHGLGSLPLRLEG 488
DB 361 SLNCKSFKFMFTKIAMDLPKFPYPPVPPAAWGEIRNVGTGLCADTKHGLGSLPLRLEG 420
QY 489 VGRGSAAMNNQVFTFWREDIRPGDPQHTKKFCFDALSHTSPTLYDCHSMKGNQLMK 548
DB 421 VGRGSAAMNNQVFTFWREDIRPGDPQHTKKFCFDALSHTSPTLYDCHSMKGNQLMK 480
QY 549 YRKDKTLHPVSGSCMDCESDHRIFMNTCNPSLTOQWLFHTNSTVLEKENRN 603
DB 481 YRKDKTLHPVSGSCMDCESDHRIFMNTCNPSLTOQWLFHTNSTVLEKENRN 535

RESULT 9

US-10-364-774-29
Sequence 29, Application US/10364774
Publication No. US20030144497A1
GENERAL INFORMATION:
APPLICANT: Donoho, Gregory
APPLICANT: Hilbun, Erin
APPLICANT: Turner, C. Alexander Jr.
APPLICANT: Friedlich, Glenn
APPLICANT: Abuin, Alejandro
APPLICANT: Zambrowicz, Brian
APPLICANT: Sands, Arthur T.
APPLICANT: Walke, D. Wade
APPLICANT: Wiganowski, Nathaniel L.
APPLICANT: Hu, Yi
APPLICANT: Kleke, James Alvin
APPLICANT: Potter, David George
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
FILE REFERENCE: LEX-0144-USA
CURRENT APPLICATION NUMBER: US/10/364,774
CURRENT FILING DATE: 2003-02-11
PRIOR APPLICATION NUMBER: US/09/795,926
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 60/185,920
PRIOR FILING DATE: 2000-02-29
PRIOR APPLICATION NUMBER: US 60/186,558
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: US 60/191,849
PRIOR FILING DATE: 2000-03-24
NUMBER OF SEQ ID NOS: 47
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 29
LENGTH: 535
TYPE: PRT
ORGANISM: homo sapiens
US-10-364-774-29

Query Match 83.8%; Score 2746.5; DB 14; Length 535;
Best Local Similarity 94.6%; Pred. No. 2.8e-258;
Matches 506; Conservative 0; Mismatches 0; Indels 29; Gaps 1;
QY 98 MDAERVDQAYRNGRIYVSDKISLNRSLPIRHRNCKSKRYLETLPNTSIIPPHNAG 157
DB 1 MDAERVDQAYRNGRIYVSDKISLNRSLPIRHRNCKSKRYLETLPNTSIIPPHNAG 60

QY 158 WSSLRTVHSVLNRSPEPELVAEIVLVDFSDREHLKKPLEDYMALPPSVRIIRTKREG 217
DB 61 WSSLRTVHSVLNRSPEPELVAEIVLVDFSDREHLKKPLEDYMALPPSVRIIRTKREG 120
QY 218 IRTMLGASVATGDTVTFLDISHCEANVMWLPLLDRIARNRTIYCPMIDVIDHDDFRYE 277
DB 121 IRTMLGASVATGDTVTFLDISHCEANVMWLPLLDRIARNRTIYCPMIDVIDHDDFRYE 180
QY 278 TOAGDMRGAFDWMYKRIPIPELOKADPSDPFSPVAGGLPAVDRKMFELGGYR 337
DB 181 TOAGDMRGAFDWMYKRIPIPELOKADPSDPFSPVAGGLPAVDRKMFELGGYR 240
QY 338 GLEIWGEQEYISFK-----VMMCGRMEDIPCSRY 368
DB 241 GLEIWGEQEYISFKGLHMLPRLVSNWPQAVFLPAPNMLAQVMMCGRMEDIPCSRY 300
QY 369 GHYRKYVPYKYPAGVSLARNLKRVAEVMWDEYAETIYORRPEYRHLASGDAVVOCKLRS 428
DB 301 GHYRKYVPYKYPAGVSLARNLKRVAEVMWDEYAETIYORRPEYRHLASGDAVVOCKLRS 360
QY 429 SLNCKSFKFMFTKIAMDLPKFPYPPVPPAAWGEIRNVGTGLCADTKHGLGSLPLRLEG 488
DB 361 SLNCKSFKFMFTKIAMDLPKFPYPPVPPAAWGEIRNVGTGLCADTKHGLGSLPLRLEG 420
QY 489 VGRGSAAMNNQVFTFWREDIRPGDPQHTKKFCFDALSHTSPTLYDCHSMKGNQLMK 548
DB 421 VGRGSAAMNNQVFTFWREDIRPGDPQHTKKFCFDALSHTSPTLYDCHSMKGNQLMK 480
QY 549 YRKDKTLHPVSGSCMDCESDHRIFMNTCNPSLTOQWLFHTNSTVLEKENRN 603
DB 481 YRKDKTLHPVSGSCMDCESDHRIFMNTCNPSLTOQWLFHTNSTVLEKENRN 535

RESULT 10

US-09-795-926-39
Sequence 39, Application US/09795926
Patent No. US20020098486A1
GENERAL INFORMATION:
APPLICANT: Donoho, Gregory
APPLICANT: Hilbun, Erin
APPLICANT: Turner, C. Alexander Jr.
APPLICANT: Friedlich, Glenn
APPLICANT: Abuin, Alejandro
APPLICANT: Zambrowicz, Brian
APPLICANT: Sands, Arthur T.
APPLICANT: Walke, D. Wade
APPLICANT: Wiganowski, Nathaniel L.
APPLICANT: Hu, Yi
APPLICANT: Kleke, James Alvin
APPLICANT: Potter, David George
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
FILE REFERENCE: LEX-0144-USA
CURRENT APPLICATION NUMBER: US/09/795,926
CURRENT FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 60/185,920
PRIOR FILING DATE: 2000-02-29
PRIOR APPLICATION NUMBER: US 60/186,558
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: US 60/191,849
PRIOR FILING DATE: 2000-03-24
NUMBER OF SEQ ID NOS: 47
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 39
LENGTH: 366
TYPE: PRT
ORGANISM: homo sapiens
US-09-795-926-39

Query Match 57.3%; Score 1877; DB 9; Length 366;
Best Local Similarity 100.0%; Pred. No. 9.6e-174;
Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKRKRLQAVLVLAALVLLPVNGMALYREROPDGTGGSGAAYAPAAAGGSHSRK 60
DB 1 MRKRKRLQAVLVLAALVLLPVNGMALYREROPDGTGGSGAAYAPAAAGGSHSRK 60
QY 61 KTFEFLDGGQKLDKMDHDEAARRDAQVNGEQRPYMTDAERVDAQYRENGFNIVYSDK 120
DB 61 KTFEFLDGGQKLDKMDHDEAARRDAQVNGEQRPYMTDAERVDAQYRENGFNIVYSDK 120
QY 121 ISLNRSLPDIRHPNCSKRYLETLPNTSIIIPHNBSGSSLLRTVSHVLRSPPELVAE1 180
DB 121 ISLNRSLPDIRHPNCSKRYLETLPNTSIIIPHNBSGSSLLRTVSHVLRSPPELVAE1 180
QY 181 VLVDDESDREHLKKPLLEDVMALEPSVRIILRTKKREGILITRMGASVATGDTVITFLDSHC 240
DB 181 VLVDDESDREHLKKPLLEDVMALEPSVRIILRTKKREGILITRMGASVATGDTVITFLDSHC 240
QY 241 EAVVNMPLPDLRIARNRKTIIVCPMIDVIDHDDFRYETOAGDAMRGAFDWMYKRIPIP 300
DB 241 EAVVNMPLPDLRIARNRKTIIVCPMIDVIDHDDFRYETOAGDAMRGAFDWMYKRIPIP 300
QY 301 PELQKADPSDPFESSPVMAAGLFVADRKFWELGSDYDGLIHWGEQYEISFKV 353
DB 301 PELQKADPSDPFESSPVMAAGLFVADRKFWELGSDYDGLIHWGEQYEISFKV 353

RESULT 11
US-10-364-774-39
Sequence 39, Application US/10364774
Publication No. US20030144497A1
GENERAL INFORMATION:
APPLICANT: Donoho, Gregory
APPLICANT: Hilbun, Erin
APPLICANT: Turner, C. Alexander Jr.
APPLICANT: Friedrich, Glenn
APPLICANT: Abuhin, Alejandro
APPLICANT: Zambrowicz, Brian
APPLICANT: Sando, Arthur T.
APPLICANT: Walke, D. Wade
APPLICANT: Wiganowski, Nathaniel L.
APPLICANT: Hu, Yi
APPLICANT: Kieke, James Alvin
APPLICANT: Potter, David George
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
FILE REFERENCE: LEX-0144-USA
CURRENT APPLICATION NUMBER: US/10/364, 774
PRIOR FILING DATE: 2003-02-11
PRIOR APPLICATION NUMBER: US/09/795, 926
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 60/185, 920
PRIOR FILING DATE: 2000-02-29
PRIOR APPLICATION NUMBER: US 60/186, 558
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: US 60/191, 849
PRIOR FILING DATE: 2000-03-24
NUMBER OF SEQ ID NOS: 47
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 39
LENGTH: 366
TYPE: PRT
ORGANISM: homo sapiens
US-10-364-774-39

Query Match 57.3%; Score 1877; DB 14; Length 366;
Best Local Similarity 100.0%; Pred. No. 9, 6e-174;
Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKRKRLQAVLVLAALVLLPVNGMALYREROPDGTGGSGAAYAPAAAGGSHSRK 60
DB 1 MRKRKRLQAVLVLAALVLLPVNGMALYREROPDGTGGSGAAYAPAAAGGSHSRK 60
QY 61 KTFEFLDGGQKLDKMDHDEAARRDAQVNGEQRPYMTDAERVDAQYRENGFNIVYSDK 120

DB 61 KTFEFLDGGQKLDKMDHDEAARRDAQVNGEQRPYMTDAERVDAQYRENGFNIVYSDK 120
QY 121 ISLNRSLPDIRHPNCSKRYLETLPNTSIIIPHNBSGSSLLRTVSHVLRSPPELVAE1 180
DB 121 ISLNRSLPDIRHPNCSKRYLETLPNTSIIIPHNBSGSSLLRTVSHVLRSPPELVAE1 180
QY 181 VLVDDESDREHLKKPLLEDVMALEPSVRIILRTKKREGILITRMGASVATGDTVITFLDSHC 240
DB 181 VLVDDESDREHLKKPLLEDVMALEPSVRIILRTKKREGILITRMGASVATGDTVITFLDSHC 240
QY 241 EAVVNMPLPDLRIARNRKTIIVCPMIDVIDHDDFRYETOAGDAMRGAFDWMYKRIPIP 300
DB 241 EAVVNMPLPDLRIARNRKTIIVCPMIDVIDHDDFRYETOAGDAMRGAFDWMYKRIPIP 300
QY 301 PELQKADPSDPFESSPVMAAGLFVADRKFWELGSDYDGLIHWGEQYEISFKV 353
DB 301 PELQKADPSDPFESSPVMAAGLFVADRKFWELGSDYDGLIHWGEQYEISFKV 353

RESULT 12
US-10-292-896-3
Sequence 3, Application US/10292896
Publication No. US20030186850A1
GENERAL INFORMATION:
APPLICANT: HASSAN, Helle
APPLICANT: REIS, Celso A.
APPLICANT: BENNETT, Eric P.
APPLICANT: CLAUSEN, Henrik
TITLE OF INVENTION: METHODS TO IDENTIFY AGENTS MODULATING FUNCTIONS OF POLYPEPTIDE G
TITLE OF INVENTION: TRANSFERASES, PHARMACEUTICAL COMPOSITIONS COMPRISING SUCH AGENT
FILE REFERENCE: 4305/1H154-US3
CURRENT APPLICATION NUMBER: US/10/292, 896
PRIOR FILING DATE: 2003-03-17
PRIOR APPLICATION NUMBER: US 60/425, 204
PRIOR FILING DATE: 2002-11-08
PRIOR APPLICATION NUMBER: PCT/DK01/00328
PRIOR FILING DATE: 2001-05-10
PRIOR APPLICATION NUMBER: US 60/203, 331
PRIOR FILING DATE: 2000-05-11
NUMBER OF SEQ ID NOS: 130
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3
LENGTH: 276
TYPE: PRT
ORGANISM: Homo sapiens
US-10-292-896-3

Query Match 42.5%; Score 1392; DB 14; Length 276;
Best Local Similarity 97.3%; Pred. No. 1, 2e-126;
Matches 250; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY 349 ISFKWMCGRMEDICSRVGHITRYKVPYKVPAGVSLA--RNLKVAEYVMDVAYEYIY 406
DB 20 LAMQVMCGRMEDICSRVGHITRYKVPYKVPAGVSLARVRLTKVAEYVMDVAYEYIY 79
QY 407 QRRPEYRHLSAGDAVAVQKLRSSLNCKSPKFWMTKLIAMDLPKYPPVEPAAAMGEIRNV 466
DB 80 QRRPEYRHLSAGDAVAVQKLRSSLNCKSPKFWMTKLIAMDLPKYPPVEPAAAMGEIRNV 139
QY 467 GTGLCADTGGAGSLPLRLGCGYRGRGEAAMNMVFTFTWRDIPRGPQHTKXCFDA 526
DB 140 GTGLCADTGGAGSLPLRLGCGYRGRGEAAMNMVFTFTWRDIPRGPQHTKXCFDA 199
QY 527 ISHTSPVTLVDCHSMKGNQLMKRYKDTLYHPVSGSCMDCSBDHIFMWTCPSSLTQ 586
DB 200 ISHTSPVTLVDCHSMKGNQLMKRYKDTLYHPVSGSCMDCSBDHIFMWTCPSSLTQ 259
QY 587 WLFPHNTSVLEKFNEN 603
DB 260 WLFPHNTSVLEKFNEN 276


```

RESULT 13
US-09-795-926-27
; Sequence 27, Application US/09795926
; Patent No. US20020098486A1
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 269
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-795-926-27

Query Match      41.8%; Score 1370; DB 9; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.6e-124;
Matches 256; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 MTDAAERVDQAYRNGENGVIVSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPPHNEG 157
DB 1 MTDAAERVDQAYRNGENGVIVSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPPHNEG 60
QY 158 WSSLRTVHSVLRSPPELVAAIYLVDDPSDRHLKKPLEDYALFSPVRIILRTKKREG 217
DB 61 WSSLRTVHSVLRSPPELVAAIYLVDDPSDRHLKKPLEDYALFSPVRIILRTKKREG 120
QY 218 IRTMLGASVATGDTTFFDSHCEANVMWLPLLDRIAARRKTIVCPMIDVIDHDDFRYE 277
DB 121 IRTMLGASVATGDTTFFDSHCEANVMWLPLLDRIAARRKTIVCPMIDVIDHDDFRYE 180
QY 278 TQAGDMRGAFFWEMYKRIPIPELOKADPSDPFESPVMAGGLFAVDRKMFWELOGYDP 337
DB 181 TQAGDMRGAFFWEMYKRIPIPELOKADPSDPFESPVMAGGLFAVDRKMFWELOGYDP 240
QY 338 GLEIWGGEQYEISFKV 353
DB 241 GLEIWGGEQYEISFKV 256

```

```

RESULT 14
US-10-364-774-27
; Sequence 27, Application US/10364774
; Patent No. US20030144497A1
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 269
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-364-774-27

Query Match      41.8%; Score 1370; DB 14; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.6e-124;
Matches 256; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 MTDAAERVDQAYRNGENGVIVSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPPHNEG 157
DB 1 MTDAAERVDQAYRNGENGVIVSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPPHNEG 60
QY 158 WSSLRTVHSVLRSPPELVAAIYLVDDPSDRHLKKPLEDYALFSPVRIILRTKKREG 217
DB 61 WSSLRTVHSVLRSPPELVAAIYLVDDPSDRHLKKPLEDYALFSPVRIILRTKKREG 120
QY 218 IRTMLGASVATGDTTFFDSHCEANVMWLPLLDRIAARRKTIVCPMIDVIDHDDFRYE 277
DB 121 IRTMLGASVATGDTTFFDSHCEANVMWLPLLDRIAARRKTIVCPMIDVIDHDDFRYE 180
QY 278 TQAGDMRGAFFWEMYKRIPIPELOKADPSDPFESPVMAGGLFAVDRKMFWELOGYDP 337
DB 181 TQAGDMRGAFFWEMYKRIPIPELOKADPSDPFESPVMAGGLFAVDRKMFWELOGYDP 240
QY 338 GLEIWGGEQYEISFKV 353
DB 241 GLEIWGGEQYEISFKV 256

```

```

RESULT 15
US-09-795-926-35
; Sequence 35, Application US/09795926
; Patent No. US20020098486A1
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926

```


/ CURRENT FILING DATE: 2001-02-28
/ PRIOR APPLICATION NUMBER: US 60/185,920
/ PRIOR FILING DATE: 2000-02-29
/ PRIOR APPLICATION NUMBER: US 60/186,558
/ PRIOR FILING DATE: 2000-03-02
/ PRIOR APPLICATION NUMBER: US 60/191,849
/ PRIOR FILING DATE: 2000-03-24
/ NUMBER OF SEQ ID NOS: 47
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 35
/ LENGTH: 321
/ TYPE: PRT
/ ORGANISM: homo sapiens
US-09-795-926-35

Query Match 40.0%; Score 1311; DB 9; Length 321;
Best Local Similarity 99.2%; Pred. No. 1.2e-118;
Matches 251; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MRRKKRLQAVLVLAALVLLPNVGLMALYREROPDGTGGGGAAPAAAGGSHSRQK 60
Db 1 MRRKKRLQAVLVLAALVLLPNVGLMALYREROPDGTGGGGAAPAAAGGSHSRQK 60
Qy 61 KTFPLDGGQKLKQWHDKEAIRDQQRVNGEQGRPYMTDAERVDQAYRENGFNIVYSDK 120
Db 61 KTFPLDGGQKLKQWHDKEAIRDQQRVNGEQGRPYMTDAERVDQAYRENGFNIVYSDK 120
Qy 121 ISLNRSIPDIRHPCNSKRYLETLPNTSIIIPHNCGWSSLRTVHSLNRSPPELVAET 180
Db 121 ISLNRSIPDIRHPCNSKRYLETLPNTSIIIPHNCGWSSLRTVHSLNRSPPELVAET 180
Qy 181 VLVDDFSDDREHLKKPLEDYVALFPSSVRLKTKKREGILRTMLGASVATGDIVITPLDSHC 240
Db 181 VLVDDFSDDREHLKKPLEDYVALFPSSVRLKTKKREGILRTMLGASVATGDIVITPLDSHC 240
Qy 241 EAVVNMVLPPLDR 253
Db 241 EAVVNMVLPPLGK 253

Search completed: December 20, 2004, 14:16:19
Job time : 151 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2004, 14:03:56 / Search time 42 Seconds
(without alignments)
1381.398 Million cell updates/sec

Title: US-10-001-851-2

Perfect score: 3278

Sequence: 1 MRKERRLQAVLVLAALV.....TQOMLFHTNISTYLEKFNEN 603

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1392.5	42.5	684	2 T26930	hypothetical prote
2	1392	42.5	276	2 T12552	hypothetical prote
3	1125	34.3	559	2 UC4223	polypeptide N-acet
4	1115	34.0	559	2 A45987	polypeptide N-acet
5	1113	34.0	601	2 T42251	polypeptide N-acet
6	1083	33.0	624	2 T42247	polypeptide N-acet
7	1079	32.9	626	2 T42246	polypeptide N-acet
8	1078.5	32.9	623	2 T42245	probable polypept
9	988.5	30.2	563	2 A88515	polypeptide N-acet
10	988.5	30.2	612	2 T42243	probable polypept
11	968	29.5	617	2 T42249	polypeptide N-acet
12	949	29.0	571	2 I37405	polypeptide N-acet
13	948.5	28.9	618	2 T42248	polypeptide N-acet
14	938	28.6	562	2 T42250	polypeptide N-acet
15	934	28.5	633	2 UC5247	polypeptide N-acet
16	906	27.6	579	2 T31549	polypeptide N-acet
17	883	26.9	589	2 T42244	probable polypept
18	801	24.4	605	2 T27397	hypothetical prote
19	632	19.3	421	2 T42252	polypeptide N-acet
20	147	4.5	1044	2 H97186	glycosyltransferas
21	139	4.2	306	2 D87531	glycosyl transfera
22	138.5	4.2	308	2 H87306	glycosyl transfera
23	136	4.1	328	2 B84263	glycosyl transfera
24	135.5	4.1	312	2 S74669	hypothetical prote
25	135.5	4.1	318	2 D87506	glycosyl transfera
26	130	4.0	322	2 AC2023	hypothetical prote
27	130	4.0	787	2 H98163	hypothetical prote
28	128	3.9	470	2 C70641	hypothetical prote
29	126	3.8	313	2 A12404	hypothetical prote

30	124	3.8	362	2 G75191	dolichol-phosphate
31	124	3.8	972	2 T09595	glucuronosyltransf
32	123	3.8	334	1 G71153	hypothetical prote
33	122	3.7	316	2 AB2868	UDP-hexose transfe
34	122	3.7	316	2 E97844	UDP-hexose transfe
35	120	3.7	260	2 E90984	probable glycosyl
36	120	3.7	260	2 H85829	glycosyl transfera
37	119.5	3.6	416	2 AE1499	conserved hypochet
38	118.5	3.6	298	2 B75096	glycosyl transfera
39	118.5	3.6	416	2 AB1141	probable glucosami
40	117.5	3.6	245	2 D87307	glycosyl transfera
41	117	3.6	343	2 A12091	glucosyltransferas
42	116	3.5	251	2 A12106	hypothetical prote
43	115.5	3.5	269	2 A13123	glycosyltransferas
44	115.5	3.5	392	2 H69814	hypothetical prote
45	115.5	3.5	477	1 US0589	endo-1,4-beta-xyla

ALIGNMENTS

RESULT 1

T26930
hypothetical protein Y45P10D.3 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #ext_change 09-Jul-2004
C/Accession: T26930
R/Mutray, A.
submitted to the EMBL Data Library, January 1998
A/Reference number: Z20288
A/Accession: T26930
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-684 <WIL>
A/Cross-references: UNIPROT:O45947; EMBL:AL021492; PIDD:CAAL6378.1; GSPDB:GN00022; CESP
A/Experimental source: clone Y45P10D
C/Genetic:
A/Gene: CESP:Y45P10D.3
A/Map position: 4 112/1, 142/3, 178/3, 230/3, 289/2, 611/2
A/Intons: 61/3, 112/1, 142/3, 178/3, 230/3, 289/2, 611/2
C/Superfamily: polypeptide N-acetylglucosaminyltransferase

Query Match	42.5%	Score 1392.5	DB 2	Length 684
Best Local Similarity	48.8%	Pred. No. 1.6e-104		
Matches 279	Conservative 87	Mismatches 175	Indels 31	Gaps 14
QY	49	PAAGGSHSRQKKTFPLGDKLKDWHDKAIRDQAQVNGEGRPYPTDAERYDQ--	106	
DB	121	PPAALGDEALDPEKRYRGH-EKIK-WEDBAAYEKERKRGPGWGRVPLPBDKEVEKEA	178	
QY	107	--AYRNGRNIVYSDKISLRSLPDRHNCNSKRYLELPNTSTIIIPFNCGSSLLRT	164	
DB	179	LSLYKANGYNAVYISDMISINRSIKDIRHKECKMMMSAKLPVSVLPFHEBNSSTLLRS	238	
QY	165	VHSLNRSPPELVAELVLVDPSDREHLKKPLLEDVY---ALPSPVILRTKKRGGILRT	221	
DB	239	VSVINRSPPELVAELVLVDPSSEKALRQPLEDFLKKKKIDHIVLVLTAKRGGILRGR	298	
QY	222	MGASVATGDVITFLDHSCEANVMPLPLDIRIARNKTIIVCPMDIVDHDPRYETQAG	281	
DB	299	QLGADATGIBILFLFHAHSEANVMPLPLDPIADBYRIVGCFVVIDCETVEYAPQ-D	357	
QY	282	DAMRGAFDWMETKRIPIPELQKADSPDPSPVMAAGLFAVDKRWFEELGSDGLBI	341	
DB	358	EGARGSPDMAFNVKRPLTRK--DRBSPTPFPSPVMAAGYFAISAKMFELGGYDGLDI	416	
QY	342	MGGEQYEISPKWMGCGRMEDIPCSRVGHYR-KYVPYK-VVAGVSLAANLKRVAVWMD	399	
DB	417	WGGEYELSEFKWQCGRNVDAPCSVAHITRKTAPEFQAAGMDVSRVRYKRVAVWMD	476	
QY	400	EVAEYTYQRPEYRHLSADVAVQKTLRSSLNCKSEKFWMTKIAMDLPRTYPPVEPAA	459	
DB	477	DYKETLYKRRPVGNNADGDLKLMKGIKREKLCCKSPDMWKEIAFPDQKXYPAVEPKASA	536	

QY 460 MGEIRVGTGLCADTRKHALGSLRLGCVR-----GRGEAAMNNMVFPTWRDIRPGD 515
 |||||
 DB 537 EGEIRVGTGFCIDTQKEQNGRFGRLKCTSDKDGGE-----QDLRLTRHDIRP-- 588
 |||||
 QY 516 PQHTKFCFPAIAHT--SPPTLYDCHSMKGNQIMKTR-KKXTLYHPVSGSCMDCSBDHR 572
 |||||
 DB 589 --KGRICFCDCSTSVKAPYLTPDCHSMKGNQIMKTRVAKQIYHPISGCLTADENGKG 646
 |||||
 QY 573 -FMNTCNPSLTLQOVLFEHTNSTVLEKFN 603
 |||||
 DB 647 FLHMKKCDSSDLQKAMQTVNDELLETROAN 678
 |||||

RESULT 2
 T12552
 hypothetical protein DKFZp586h0623.1 - human (fragments)
 C/Species: Homo sapiens (man)
 C/Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 09-Jul-2004
 C/Accession: T12552
 R/Author: W.; Winkler, U.; Mewes, H.W.; Gaassenhuber, J.; Wiemann, S.
 Submitted to the Protein Sequence Database, June 1999
 A/Reference number: Z17527
 A/Accession: T12552
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-150/151-276 <ANS>
 A/Experimental source: UNIPROT:Q9Y4M4; EMBL:AL096739
 A/Note: the cDNA sequence contains a -1 frameshift near codon 150
 C/Genetics:
 A/Note: DKFZp586h0623.1
 C/Superfamily: polypeptide N-acetylglucosaminyltransferase

Query Match 42.5%; Score 1392; DB 2; Length 276;
 Best Local Similarity 97.3%; Pred. No. 4,6e-105;
 Matches 250; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY 349 ISFKVWCGRMEDICSRVGHYRKVPYKPVAGVSLA--RNILKVAEVMDEYAEYI 406
 |||||
 DB 20 LAMQVWCGRMEDICSRVGHYRKVPYKPVAGVSLA--RNILKVAEVMDEYAEYI 79
 |||||
 QY 407 QRRPEYRHLSAGVAVQKIRSSSLNCKSFEMPTKIAMDLPKYPPVEPPAAWGEIRNY 466
 |||||
 DB 80 QRRPEYRHLSAGVAVQKIRSSSLNCKSFEMPTKIAMDLPKYPPVEPPAAWGEIRNY 139
 |||||
 QY 467 GTGLCADTRKHALGSLRLGCVRGEAAMNNMVFPTWRDIRPGDQHTKFCFPA 526
 |||||
 DB 140 GTGLCADTRKHALGSLRLGCVRGEAAMNNMVFPTWRDIRPGDQHTKFCFPA 199
 |||||
 QY 527 ISHTSPVTLVDCHSMKGNQIMKTRKXTLYHPVSGSCMDCSBDHRIFMNTCNPSLTLQ 586
 |||||
 DB 200 ISHTSPVTLVDCHSMKGNQIMKTRKXTLYHPVSGSCMDCSBDHRIFMNTCNPSLTLQ 259
 |||||
 QY 587 WLFHTNSTVLEKFN 603
 |||||
 DB 260 WLFHTNSTVLEKFN 276
 |||||

RESULT 3
 JC4223
 polypeptide N-acetylglucosaminyltransferase (EC 2.4.1.41) - human
 N/Alternate names: GalNAC-transferase
 C/Species: Homo sapiens (man)
 C/Date: 14-Nov-1995 #sequence_revision 08-Feb-1996 #text_change 09-Jul-2004
 R/Author: J.A.; Maylor, J.N.; Baker, C.A.; Thomsen, D.R.; Homa, F.L.; Elhammer, A.P.
 J. Biochem. 118, 568-574, 1995
 A/Title: cDNA cloning, expression, and chromosomal localization of a human UDP-GalNAC:
 A/Reference number: JC4223; MUID:96115928; PMID:8690719
 A/Accession: JC4223
 A/Molecule type: mRNA
 A/Residues: 1-559 <MEU>

A/Cross-references: UNIPROT:Q10472
 A/Experimental source: salivary gland
 A/Note: The authors translated the codon AAT for residue 264 as Asp
 R/White, T.; Bennett, B.P.; Paul, E.; Takio, K.; Stenssen, T.; Bonding, N.; Clausen, H.
 J. Biol. Chem. 270, 24156-24165, 1995
 A/Title: Purification and cDNA cloning of a human UDP-GalNAC:polypeptide N-Acetylglucosaminyltransferase
 A/Reference number: 137404; MUID:96025800; PMID:7592619
 A/Accession: 137404
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-559 <RES>
 A/Cross-references: EMBL:X85018; NID:9971458; PIDN:CA45380.1; PID:9971459
 A/Comment: This enzyme catalyzes the initial reaction in O-linked (mucin type) oligosac
 ccharide pathway.
 C/Genetics:
 A/Map position: 18
 C/Superfamily: polypeptide N-acetylglucosaminyltransferase
 C/Keywords: Chromosomal protein; glycoprotein; glycosyltransferase; hexosyltransferase;
 F/2-28/Domain: transmembrane #status predicted <TM>
 F/29-559/Domain: endoplasmic reticulum lumenal #status predicted <LUM>
 F/35,141,541,552/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F/117,118,288/Binding site: carbohydrate (Thr) (covalent) #status predicted
 F/119/Binding site: carbohydrate (Ser) (covalent) #status predicted

Query Match 34.3%; Score 1125; DB 2; Length 559;
 Best Local Similarity 44.0%; Pred. No. 5,6e-83;
 Matches 232; Conservative 87; Mismatches 170; Indels 38; Gaps 16;

QY 88 GNGEGRP--YEMTDKERDQAVRENGRIYVSDKISLNRSLPDRIRHPCNSKRYLETIP 145
 |||||
 DB 57 GPEMGKPVVIPPEDDKKEMKFNQFNLMSSEMIALNRSLPDRIRHPCNSKRYLETIP 116
 |||||
 QY 146 NTSIIIPFNEGMSLIRTVHSLNRSPELVAEIVLVDPFDRHKKPLBDM-ALRP 204
 |||||
 DB 117 TTSVIVFVFNEMASTLIRTVHSLNRSPELVAEIVLVDPFDRHKKPLBDM-ALRP 176
 |||||
 QY 205 SVRIIRTKRSGILIRTKLGAASVAGDVITFLDSHCANVNLPLLDRIARNRKTIVCP 264
 |||||
 DB 177 PVHIVMEGRSGILIRTKLGAASVAGDVITFLDSHCANVNLPLLDRIARNRKTIVCP 236
 |||||
 QY 265 MIVDIVDHPFRFTQAG-DAMGAPDMEMTYKRIPIRP--ELQKADBPDPESPVMAG 320
 |||||
 DB 237 IIVDIVSDTFEY--MAGSDMTYGGFVWKLNFVYVPPQREMRRKGRDRLPVRTPTMAG 294
 |||||
 QY 321 LPAVDKFWFVAGYDPPGLEIENGEOYELSFVWCGRMEDICSRVGHYRKVPYK 380
 |||||
 DB 295 LPSIDNDYQELGTIDAGNDINGENLEISFIMQCGTLEIVTSGHGVHVRKATPTPF 354
 |||||
 QY 381 PAVGS--LARNILKVAEVMDEYAEYIYQRRPEYRHLSAGVAVQKIRSSSLNCKSF 438
 |||||
 DB 355 PGGTGQIINKNNRRLAEVMDFFKNFPYIISGVTVKVDYDSSRVGLRHKLQCKPFSWY 414
 |||||
 QY 439 MKRIAND--LPKFPVPEPPAAWGEIRVVGGLCADTRKHALGSLRLGCVRGEA 496
 |||||
 DB 415 LENIYPPDSQIPRHY-----PSLGEIRNVEVETNQCIDNNARKENKVGIFNC-HGNG-- 463
 |||||
 QY 497 KNNMQVFTWRDIRPGDQHTKFCFPAISHTSPVTLVDCHSMKGNQIMKTRKDX--TL 555
 |||||
 DB 464 --GNQVFSYTKANKETRTD-----LCDVSLNGLVPMYMKCHHLKGNQIMYEDPVKLT 515
 |||||

QY 556 YHPVSGSCMD-CSESDHRI-FMNTCNPSLTLQOVLFEHTNSTVLEKFN 600
 |||||
 DB 516 QVHNSQCUDKATEDSQVPSIRDCN-GSRSGQMLR--NVTLPEIF 559
 |||||

RESULT 4
 A45987
 polypeptide N-acetylglucosaminyltransferase (EC 2.4.1.41) - bovine
 C/Species: Bos primigenius taurus (cattle)
 C/Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 09-Jul-2004
 C/Accession: A45987; A48530
 R/Homa, F.L.; Hollander, T.; Lehman, D.J.; Thomsen, D.R.; Elhammer, A.P.
 J. Biol. Chem. 268, 12609-12616, 1993

Query Match	34.0%;	Score 1115;	DB 2;	Length 559;
Best Local Similarity	43.6%;	Pred. No. 3.6e-82;		
Matches 230; Conservative	89;	Mismatches 170;	Indels 38;	Gaps 16;

RESULT 5
T42251
polypeptide N-acetylglactosaminyltransferase (EC 2.4.1.41) 7 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T42251
R:Hagen, F.K.; Nehrke, K.
J. Biol. Chem. 273, 8268-8277, 1998
A:Title: cDNA cloning and expression of a family of UDP-N-acetyl-D-glactosamine:polypeptide
A:Reference number: Z22126; MUID:96192620; PMID:9525933
A:Accession: T42251
A:Status: preliminary; translated from GB/EMBL/DBJ

Query Match	34.0%;	Score 113;	DB 2;	Length 601;
Best Local Similarity	42.0%;	Pred. No. 5.8e-82;		
Matches 226;	Conservative 90;	Mismatches 186;	Indels 36;	Gaps 13;

RESULT

Query Match	33.0%;	Score 1083;	DB 2;	Length 624;
Best Local Similarity	45.0%;	Pred. No. 1.7e-79;		
Matches 226; Conservative	76;	Mismatches 166;	Indels 34;	Gaps 13;

T42249
 polypeptide N-acetylglucosaminyltransferase (EC 2.4.1.41) 6b - Caenorhabditis elegans
 C/Species: Caenorhabditis elegans
 C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
 C/Accession: T42249
 R/Hagen, F.K.; Nehrkke, K.
 J. Biol. Chem. 273, 8268-8277, 1998
 A/Title: CDNA cloning and expression of a family of UDP-N-acetyl-D-glucosamine:Polypept
 A/Reference number: 222126; MUID:98192620; PMID:9525933
 A/Accession: T42249
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-617 <HAG>
 A/Cross-references: UNIPROT:O61395; EMBL:AF031839; NID:93047198; PID:AA013675.1; PID:93
 C/Genetics:
 A/Gene: gly-6
 C/Superfamily: polypeptide N-acetylglucosaminyltransferase
 C/Keywords: glycosyltransferase; hexosyltransferase

Query Match 29.5%; Score 968; DB 2; Length 617;
 Best Local Similarity 39.3%; Pred. No. 3,4e-70;

Matches 216; Conservative 96; Mismatches 162; Indels 76; Gaps 22;

QY 73 DHWDKAIARRDAQRVNGEGRRPYMTDAERV-DQAYRENGFNIVSDKISLNRSLPDIR 131
 DB 97 DW-----GGGAGVSHLPREGQKADSTFAVQFNLVSDGISVRSLSPEIR 143
 QY 132 HPCNSKRYLETLPNTSIIPFNEGSSLRVSHVSLNRSPELVAEIVLVDPGDRH 191
 DB 144 KPSCRNMTYFDNPTTSVILVHNEAVSTLRTVWSYIDSPKLEIILVDPGDRER 203
 QY 192 LKKP-LEDYVALRPS-VRIILTKKREGLIRTRMLGASVATGDTYTFDLSHCANVWLP 249
 DB 204 LRPTDITLTKRPTDIKIRSKERVGLIARMMGAGAGDVLITDSCCECTKMWLP 263
 QY 250 LLDRIARNKRTIVCEMIDVIDHDDFRYETGAGDAMRGAPFWEYTKRIPPELQK--A 306
 DB 264 LTRIKLRKAVPCPYDIINDNTFOYQ-KGIMEFRGGFYMNLOFRWYGMPTAMAKOHL 322
 QY 307 DPEDPESPVMAGLFAVDKMFMEIAGYDGLINGBOYEISFKVMCGGMEDIPCS 366
 DB 323 DPGPIESPTMAGGLPSINNYPELGEYDPMGMDINGENLEMSFRIMQCGRVELLP 382
 QY 367 RVGHIYRKYVPYKVP--AGVSLARMLKVAEYVMDYAEYIYORRPE-VRLHSGADVAV 422
 DB 383 HGVHVRKSSPHDFPGSSGKVLNTNLRAVEYVMDKMYFYKIAPOARMSSIDVSE 442
 QY 423 QKKLRSSLNCKSKFMFTKIAMDLPKFYRPPVPPAAWGEIRVVGGLC----ADT--- 474
 DB 443 RVELRKTLNCKSKFMYLQNVFOD--HFLP---TPLDREFGMSN--SNYCTAFRPGDTGP 495
 QY 475 KHGALGSLRLBECVGRGAANNMNOVFTWRREDIRPGDPHTKKFCFGLS--HTS- 531
 DB 496 NHRLLSP-----CTMG-----FDLMQMLVYTDGRRIRIDE-----HLCLSVQLHTTS 540
 QY 532 --PVTLYDCHSMKNQMLKTR-KDKTLVHVSQSCMDCSBDRIF-----MNT 577
 DB 541 DWKIQKECGAF-DTEVWDFKPKIGRFQNRKTKGLCLASP-----ITDPTDSENPPIVQ 595
 QY 578 CNPSSLTQW 587
 DB 596 CRSSNDROW 605

RESULT 12
 T42405
 polypeptide N-acetylglucosaminyltransferase (EC 2.4.1.41) - human
 C/Species: Homo sapiens (man)
 C/Date: 24-May-1996 #sequence_revision 24-May-1996 #text_change 09-Jul-2004
 C/Accession: T42405
 R/Hagen, F.K.; Nehrkke, K.
 J. Biol. Chem. 270, 24156-24165, 1995
 A/Title: Purification and CDNA cloning of a human UDP-GalNAc:polypeptide N-Acetylglucos

A/Reference number: T42405; MUID:96025800; PMID:7592619
 A/Accession: T42405
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-571 <RHS>
 A/Cross-references: UNIPROT:Q10471; EMBL:X85019; NID:99711460; PID:CAA59381.1; PID:997114
 C/Genetics:
 A/Gene: GALNT2; GALNAC
 A/Cross-references: GDB:696223
 A/Map position: 16q24-16q24
 C/Superfamily: polypeptide N-acetylglucosaminyltransferase
 C/Keywords: glycosyltransferase; hexosyltransferase

Query Match 29.0%; Score 949; DB 2; Length 571;
 Best Local Similarity 33.5%; Pred. No. 1.1e-68;

Matches 221; Conservative 103; Mismatches 170; Indels 166; Gaps 21;

QY 1 MRRKEKLLQAVLVALLVLLPNVGMALYREORDDGTPGSGAAVAPACGSHSROK 60
 DB 1 MRRSRMLLCFAFLWVLGIA-----YYM-----SGGSLMAGAGGGA----- 39
 QY 61 KTFPLGSGKLDMDHDKAARRDAQRVNGEG-----GR-PYMTDAERV----- 104
 DB 40 -----GRK-EDWNEDIPKDKDLHNSGEEKAGSMETLPPGKRWPFNDQAYVGGTM 91
 QY 105 ----DQAYRENGFNIVSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIPFNEGWS 160
 DB 92 VNSGQDPYARNFQVQESKLMMDALIPTRIDQCRKMRVLDLPAISVITFHNBARA 151
 QY 161 LRTVSHVSLNRSPELVAEIVLVDPGDRHLKKPLED--VVALPSPVILRTKKREGI 218
 DB 152 LRTVSVLVKKSPPHLIKSIILVDDYSNDP-----EDGALLKIEKVVLNDRREGIM 205
 QY 219 KRRMKGASVATGDTYTFDLSHCANVWLPRLDIRARNTKIVCEMIDVIDHDDFRY 278
 DB 206 RSRVRGADAQAOKLTFPLDLSHCENHMLEPLERVAEDRTVSPITDIVIMDNFOYG 265
 QY 279 QAGDARAFDPEMYKRIPIPELOKA---DPSDFESPVMAGLFAVDKMFMEIAGY 335
 DB 266 ASAD-LKGFMDNVNLYKMYMPBEQRRSQGNVPVPIKTPMAGLFVWDKRYFELGK 324
 QY 336 DPGLEIWSGEQYEISFKVMCGRMEDIPCSRVGHYIKRYVPYKVP--AGVSLARMLK 393
 DB 325 DMMDDVWGENSEISFRVWQCGSLIIPCSRVGHYFRKQHPYTFPGSGGTAFANTRBA 384
 QY 394 AEVMDVEYAEIYORRPEVRHLSAGDVAAQKLRBSLNCKSKFMFTKIAMDLPKFY 453
 DB 385 AEVMDVEYKPYFAVPSARVPYQISRLRLKRLSKPKMYLENV-----YREL 437
 QY 454 EPPAAWGEIRVVGGLCADTGHGALGSPRLLEGVGRGAANNMNOVFTWRREDIRP 513
 DB 438 RVPDH-----QDIAPGL----- 450
 QY 514 GPDQHTKFCFPAISHTSP-VTLYDCHSMKNQMLKTRKDKTLVH-----PVSGS 562
 DB 451 --QGGTN--CLDTLGHFADGVGVVECHNAGNQEALTKERSYKMDCLTVVDRAPS 506
 QY 563 CM---DCSSDR-----IMNTCNPSSLTQW 589
 DB 507 LIRLQGRENDSRQKWEQIEGNSKLRLHVSNTCLDSRTAKSGGLVEYVGP-ALSOQMK 565

RESULT 13
 T42248
 polypeptide N-acetylglucosaminyltransferase (EC 2.4.1.41) - Caenorhabditis elegans
 C/Species: Caenorhabditis elegans
 C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
 C/Accession: T42248; T23138
 R/Hagen, F.K.; Nehrkke, K.
 J. Biol. Chem. 273, 8268-8277, 1998
 A/Title: CDNA cloning and expression of a family of UDP-N-acetyl-D-glucosamine:Polypept
 A/Reference number: 222126; MUID:98192620; PMID:9525933
 A/Accession: T42248

Db 122 GASGKPEFKITHLSPBEOKEKERGETYKCFNAPASDRISLHRDLGPDTRPEECIEQKFKRC 181
QY 142 ETLPTNTSIIPFNNEGSSILRTVHSLVNSRPEELVAEIVLVDQFSDBEHLKKPLSDYMA 201
Db 182 PPLPTTSVILVFNEANSITLRTVHSLVSSPAILKEITLVDDASVDDYLHEKELEBYIK 241
QY 202 LPPSVRILRTKKEGLIRTMLGASVATGDTVITFLDSHCENYVNLPPLLDRIARNRXTI 261
Db 242 QFETIVKIVRQEKGLITARLGAAVATAETLTFDLAHGECFYGMLEPILARIENYTA 301
QY 262 VCPMIDVITDHDPRYETQA--GDAMRGATDWMY--KRIPLPELOKADPSDPESPV 316
Db 302 VSPDIASIDLNTPEFNFNPPSYGSHNRGNPDWSLSFGWESLPDHEKORRDETYPIKTP 361
QY 317 MAGGLFVADRKFMELGQYDQGLIEMGEGYEISFKVMGSGRMEDIPGSRVGHYRKXV 376
Db 362 FAGGLFSISKKEIHEHIGSYDEMEIMGENIEMSFVWQCGGLEIWPSCSVGVGVFRSKS 421
QY 377 PYKVPAGVS-LAENLKRVAEVMDEYAEYIYORRPE---YRHLASAGDVAVQKKRSSLN 431
Db 422 PHTFPKGTQVIANQVRLAEVMMDEYKEIFYRNTDAKIVKQKSGDLSKREIKKRLQ 481
QY 432 CKSGKFMETKIANDLPKFPYVBPAAAMEIRNVGTGLCADT-KHGAUGSPRLBQV 490
Db 482 CKNFTWYLNIT--YPEAYVPDLNPVIS-GYIKSVGQPLCLDVGENNQGKPLILYTC-H 536
QY 491 GRGEAAMNNNVQFTFTWRREDIRPGDPQHT--KKCFDAISHTSPV--TLYDCHS--MKN 544
Db 537 GLG----GNQYFYSQREIR----HNIOKELCLHATOGVVQLKACVYKGRRTIAPGE 586
QY 545 QLMKRYRKDTLYHPVSGSCMDCSESDHRIEMNTCNPSLTQOWLF 589
Db 587 QIWEIRRDQLLVNPLFMCLTS-SNGEHPNLV-PCDATDILQKWIF 629

Search completed: December 20, 2004, 14:13:49
Job time : 49 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2004, 13:55:48 ; Search time 197 Seconds

(without alignments)
1761.172 Million cell updates/sec

Title: US-10-001-851-2

Perfect score: 3278

Sequence: 1 MRKKEKRLQNALVALVALV.....TQGWLPFHNTSTVLEKFNEN 603

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3278	100.0	603	1 GL10_HUMAN	Q86SR1 h polypepti
2	3157	96.3	603	1 GL10_HUMAN	Q92577 r polypepti
3	3138	95.7	634	2 BAD21405	Bad21405 mus muscu
4	3137	95.7	603	1 GL10_MOUSE	Q6P967 m polypepti
5	3137	95.7	603	2 AAH60617	Aah60617 mus muscu
6	1404	42.8	644	2 Q700B9	Q7Q069 anopheles g
7	1392.5	42.5	632	1 GL10_CAEEL	Q45947 caenorhabdi
8	1360	41.5	666	1 GLT6_DROME	Q6W16 drosophila
9	1360	41.5	666	2 AAQ56703	Aaqs6703 drosophila
10	1341.5	40.9	599	2 Q7P2M5	Q7Q2M5 anopheles g
11	1247	38.0	1003	2 Q7QDR0	Q7QDR0 anopheles g
12	1186	36.2	659	1 GLT4_DROME	Q81442 drosophila
13	1186	36.2	659	2 AAQ56701	Aaqs6701 drosophila
14	1172.5	35.8	650	1 GLT8_DROME	Q8MFC3 anopheles g
15	1159	35.4	645	2 Q7Q8T5	Q7Q8T5 anopheles g
16	1144	34.9	518	2 Q7Q048	Q7Q048 anopheles g
17	1125	34.3	559	1 GLT1_HUMAN	Q1Q472 h polypepti
18	1117	34.0	559	1 GLT1_PIG	Q29141 s polypepti
19	1116	34.0	559	1 GLT1_RAT	Q10473 r polypepti
20	1115	34.0	559	1 GLT1_BOVIN	Q07537 b polypepti
21	1115	34.0	559	1 GLT1_MOUSE	Q08912 m polypepti
22	1113	34.0	601	1 GLT7_CAEEL	Q6GMS1 caenorhabdi
23	1112.5	33.9	556	2 Q6GMS1	Q6GMS1 xenopus lae
24	1111.5	33.9	617	2 AAQ56702	Aaqs6702 drosophila
25	1109	33.8	556	1 GL13_MOUSE	Q8CF3 r polypepti
26	1109	33.8	556	1 GL13_RAT	Q6W29 r polypepti
27	1109	33.8	556	2 AAQ5749	Aaqs749 ratu no
28	1106	33.7	630	1 GLT5_DROME	Q6W17 drosophila
29	1106	33.7	630	1 GLT5_HUMAN	Q81UC8 h polypepti
30	1087	33.2	559	2 Q6P81	Q6P81 xenopus lae
31	1087	33.2	559	2 AAH60419	Aah60419 xenopus l

32	1083	33.0	626	1 GLT5_CAEEL	Q9521 caenorhabdi
33	1083	33.0	653	2 Q6N824	Q6N824 xenopus lae
34	1083	33.0	653	2 AAH70527	Aah70527 xenopus l
35	1053	32.1	653	2 Q6XK21	Q6XK21 xenopus tro
36	1053	32.1	653	2 AAH67317	Aah67317 xenopus t
37	1050	32.0	657	1 GLT7_HUMAN	Q8652 homo sapien
38	1049	32.0	930	1 GLT5_MOUSE	Q8C102 m polypepti
39	1047	31.9	930	1 GLT7_MOUSE	Q80V40 mus musculi
40	1040	31.7	657	1 GLT7_RAT	Q9RC05 ratu no
41	1040	31.7	930	1 GLT5_RAT	Q88422 r polypepti
42	1037	31.6	606	2 Q7QEH0	Q7QEH0 anopheles g
43	1036	31.6	578	1 GLT4_HUMAN	Q8W48 h polypepti
44	1035.5	31.6	591	1 GLT7_DROME	Q8W48 drosophila
45	1035.5	31.6	591	2 AAQ56704	Aaqs6704 drosophila

ALIGNMENTS

RESULT 1

ID	GL10_HUMAN	STANDARD	PRT	603 AA.
AC	Q86SR1; Q6INS6; Q86VP8; Q8IXJ2; Q96IV2; Q9HBE1; Q9Y4M4;			
DT	01-OCT-2004 (Rel. 45, Created)			
DT	01-OCT-2004 (Rel. 45, Last sequence update)			
DT	01-OCT-2004 (Rel. 45, Last annotation update)			
DE	Polypeptide N-acetylglucosaminyltransferase 10 (EC 2.4.1.41)			
DE	(Protein-UDP acetylglucosaminyltransferase 10) (UDP-			
DE	GalNac:polypeptide N-acetylglucosaminyltransferase 10) (Polypeptide			
DE	GalNac transferase 10) (GalNac-T10) (pp-GalNase 10).			
GN	Name=GLNT10;			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
OX	[1]			
RP	SEQUENCE FROM N.A. (ISOFORM 1), ENZYME ACTIVITY, AND TISSUE			
RP	SPECIFICITY.			
RC	TISSUE=Colon cancer;			
RX	MEDLINE=22304871; PubMed=12417297;			
RA	Cheng L., Tachibana K., Zhang Y., Guo J.-M., Tachibana K.K.,			
RA	Kameyama A., Wang H., Hiruma T., Iwasaki H., Togayachi A., Kudo T.,			
RA	Narimatsu H.;			
RT	"Characterization of a novel human UDP-GalNac transferase, pp-GalNac-			
RT	T10.";			
RL	FEBS Lett. 531:115-121 (2002).			
RN	[2]			
RP	SEQUENCE FROM N.A. (ISOFORM 1).			
RL	Bennett E.P.;			
RL	Submitted (Aug-2002) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A. (ISOFORM 4).			
RP	TISSUE=Uterus;			
RC	Ansoorge W., Wikner U., Mewes H.-W., Gassenhuber J., Wiemann S.;			
RL	Submitted (Jun-1999) to the EMBL/GenBank/DBJ databases.			
[1]				
SEQUENCE FROM N.A. (ISOFORM 5), AND SEQUENCE OF 191-603 FROM N.A.				
(ISOFORM 1).				
TISSUE=Kidney, Skin, and Testis;				
RX	MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;			
RA	Straussberg R.L., Feigold E.A., Grouse L.H., Derge J.G.,			
RA	Klausner R.D., Collins F.S., Wagner L., Siemien C.M., Schuler G.D.,			
RA	Altschul S.F., Zeeberg B., Bueltow K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Scapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Schaefer T.E.,			
RA	Bromstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,			
RA	Riba S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mulian S.J.,			
RA	Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gundarone P.H.,			
RA	Richards S.K., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,			
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,			
RA	Phley J., Helton B., Ketterman M., Madan A., Rodrigues S., Sanchez A.,			
RA	Whiting M., Madan A., Young A.C., Shcherchenko Y., Bouffard G.G.,			

RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield V.S.N., Krzywinski M.I., Skalek U., Smallos D.E.,
 RA Schenck A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RP [5]
 RP SEQUENCE OF 8-366 FROM N.A. (ISOFORM 3), AND SEQUENCE OF 117-603 FROM
 RP N.A. (ISOFORM 2).
 RC TISSUE=Placenta, and Spleen;
 RX PubMed=14702039; DOI=10.1038/ng1285;
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
 RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
 RA Sekine M., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
 RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
 RA Nagahashi K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
 RA Shiratori A., Sudo H., Hosoki T., Kaku Y., Kodaira H., Kondo H.,
 RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
 RA Omura Y., Abe K., Kamihara K., Katsuma N., Sato K., Tanikawa M.,
 RA Yamazaki M., Niimiyama K., Ishibashi T., Yamashita H., Murakawa K.,
 RA Fujimori K., Tanai H., Kimura M., Watanabe M., Hirakawa S., Chiba Y.,
 RA Iehida S., Ono Y., Takiguchi S., Watanabe S., Yoshida M., Hattori T.,
 RA Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase I.-O.,
 RA Nomura Y., Togiyama S., Komai F., Hara R., Takeuchi K., Arita M.,
 RA Imose N., Musashino K., Yuki H., Oshima A., Sasaki N., Aotsuka S.,
 RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,
 RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakabe H.,
 RA Hibiagi H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
 RA Yamazaki M., Komiyama M., Tashiro H., Tanigami A., Fujitani Y.,
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiraio M., Omori Y.,
 RA Kawabata A., Hikiji T., Kobayashi N., Inagaki H., Ikema Y., Okamoto S.,
 RA Ohtsuki R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
 RA Matsunaga K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
 RA Togaishi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Maeno Y., Yamashita R.,
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sigano S.;
 RT "Complete sequencing and characterization of 21,243 full-length human
 RT cDNAs.";
 RL Nat. Genet. 36:40-45(2004).
 CC -1- FUNCTION: Catalyzes the initial reaction in O-linked
 CC oligosaccharide biosynthesis, the transfer of an N-acetyl-D-
 CC galactosamine residue to a serine or threonine residue on the
 CC protein receptor. Has activity toward Muc5Ac and E2 peptide
 CC substrates.
 CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =
 CC UDP + N-acetyl-D-galactosaminyl-polypeptide.
 CC -1- COFACTOR: Manganese and calcium (By similarity).
 CC -1- PATHWAY: Glycosylation.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By
 CC similarity).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=5;
 CC Name=1;
 CC IsoId=Q86SR1-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q86SR1-2; Sequence=VSP_011209;
 CC Note=No experimental confirmation available;
 CC Name=3;
 CC IsoId=Q86SR1-3; Sequence=VSP_011212; VSP_011213;
 CC Note=No experimental confirmation available;
 CC Name=4;
 CC IsoId=Q86SR1-4; Sequence=VSP_011207; VSP_011208; VSP_011214;
 CC Note=No experimental confirmation available;
 CC Name=5;
 CC IsoId=Q86SR1-5; Sequence=VSP_011210; VSP_011211;
 CC Note=No experimental confirmation available;
 CC -1- TISSUE SPECIFICITY: Widely expressed. Expressed at high level in
 CC small intestine, and at intermediate levels in stomach, pancreas,
 CC ovary, thyroid gland and spleen. Weakly expressed in other

CC tissues.
 CC -1- DOMAIN: There are two conserved domains in the glycosyltransferase
 CC region: the N-terminal domain (domain A, also called GT1 motif),
 CC which is probably involved in manganese coordination and substrate
 CC binding and the C-terminal domain (domain B, also called
 CC Gal/GalNAc-T motif), which is probably involved in catalytic
 CC reaction and UDP-Gal binding (By similarity).
 CC -1- DOMAIN: The ricin B-type lectin domain binds to GalNAc and
 CC contributes to the glycospecificity (By similarity).
 CC -1- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T
 CC subfamily.
 CC -1- SIMILARITY: Contains 1 ricin B-type lectin domain.
 CC -1- CAUTION: According to experiments made in rat, this enzyme is
 CC unable to transfer GalNAc onto serine or threonine residue on the
 CC protein receptor, but instead requires the prior addition of a
 CC GalNAc on a peptide before adding additional GalNAc moieties,
 CC thereby acting as a glycopeptide transferase.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 CC EMBL; AB078145; BAC56890.1; -
 CC EMBL; AJ505950; CAB44532.1; -
 CC EMBL; AI096739; CAB445378.1; -
 CC EMBL; BC007224; AAH07224.2; -
 CC EMBL; BC050333; AAH50333.1; -
 CC EMBL; BC072450; AAH72450.1; -
 CC EMBL; AK023782; BAB14676.1; -
 CC EMBL; AK074132; BAB84958.1; -
 CC PIR; T12552; T12552.
 CC HSSP; P26514; IKM1.
 CC GeneW; HGNC:19873; GALNT10.
 CC MIM; 608043; -
 CC InterPro; IPR001173; Glyco trans 2.
 CC InterPro; IPR008997; RicinB like-
 CC InterPro; IPR000772; RicinB lectin.
 CC Pfam; PF00535; Glycoe trans_2; 1.
 CC Pfam; PF00652; Ricin_B lectin; 3.
 CC SMART; SM00458; Ricin; 1.
 CC PROSITE; PS50231; RICIN B LECTIN; 1.
 CC KEGG; Alternative splicing; Calcium; Glycosyltransferase; Golgi stack;
 CC lectin; Manganese; Signal-anchor; Transferase; Transmembrane.
 CC FT DOMAIN 1 11
 CC TRANSMEM 12 31
 CC FT DOMAIN 32 603
 CC FT DOMAIN 144 253
 CC FT DOMAIN 311 373
 CC FT DOMAIN 458 590
 CC FT DISULFID 471 488
 CC FT DISULFID 523 538
 CC FT DISULFID 563 578
 CC FT CARBOHYD 124 146
 CC FT CARBOHYD 146 166
 CC FT CARBOHYD 593 593
 CC FT CARSPPLIC 1 329
 CC FT VARSPLIC 330 352
 CC FT VARSPLIC 190 251
 CC FT VARSPLIC 190 202
 CC FT VARSPLIC 203 603
 CC FT VARSPLIC 354 366
 CC -----
 CC /FTid=VSP_011208.
 CC Missing (in isoform 2).
 CC /FTid=VSP_011209.
 CC EHLKPLBDYML -> DLPASTPSPFVC (in
 CC isoform 5).
 CC /FTid=VSP_011210.
 CC Missing (in isoform 5).
 CC /FTid=VSP_011211.
 CC NMCGRMEDIPCS -> SOLSRPVLGTAS (in

Query Match 96.3%; Score 3157; DB 1; Length 603;
 Best Local Similarity 96.0%; Pred. No. 2e-247;
 Matches 579; Conservative 7; Mismatches 17; Indels 0; Gaps 0;

QY 1 MRKEKRLLOAVLVAALVLLPNVGLMALYERQPDGTGGGAAPAAAGGSHSRK 60
 DB 1 MRKEKRLLOAVLVAALVLLPNVGLMALYERQPDGTGGGAAPAAAGGSHSRK 60
 QY 61 KTFPLGDGQKLDKMDHKEAIRDAQVNGEGQRPVPMTDABRVDAVRENGFNIVSDK 120
 DB 61 KTFPLGAEQRLKDMHKEAIRDAQVNGEGQRPVPMTDABRVDAVRENGFNIVSDK 120
 QY 121 ISLNRSLPDIRHPNCNKRYLETLPNTSIIIPHNEGSSLLRTVSVLNRSPELVAEI 180
 DB 121 ISLNRSLPDIRHPNCNKRYLETLPNTSIIIPHNEGSSLLRTVSVLNRSPELVAEI 180
 QY 181 VLVDDSDREHLKKPLEDYMALPPSVRIILRTKREGILRTMLGASAAATGDTITFLDSHC 240
 DB 181 VLVDDSDREHLKKPLEDYMALPPSVRIILRTKREGILRTMLGASAAATGDTITFLDSHC 240
 QY 241 EAVNMVLPPLDLRIANRRTIVCPMIDVIDHDPRYETQAGDAMRGAFPMEMYKRIPI 300
 DB 241 EAVNMVLPPLDLRIANRRTIVCPMIDVIDHDPRYETQAGDAMRGAFPMEMYKRIPI 300
 QY 301 PELQKADPSDPESPVMAGGLFVADRKKFWELGSDYDGLGIEWGEGYEISFKVMCGGRM 360
 DB 301 PELQKADPSDPESPVMAGGLFVADRKKFWELGSDYDGLGIEWGEGYEISFKVMCGGRM 360
 QY 361 EDIPCSRVGHIYRKYPYVPAGVSLARNLKRVAEVMDEYAEYIQRREYRHLSAGDV 420
 DB 361 EDIPCSRVGHIYRKYPYVPAGVSLARNLKRVAEVMDEYAEYIQRREYRHLSAGDV 420
 QY 421 AVQKRLRSSLNCKSFPMFTKIAMDLPKFYPPVEPPAAWGIIRNVTGLCADTKGALG 480
 DB 421 AVQKRLRSSLNCKSFPMFTKIAMDLPKFYPPVEPPAAWGIIRNVTGLCADTKGALG 480
 QY 481 SPRLLETCIRGGEAAMNMQVFTFTWRREDIRPGDPQHTKKCFDAVSHTSPTLYDCHS 540
 DB 481 SPRLLETCIRGGEAAMNMQVFTFTWRREDIRPGDPQHTKKCFDAVSHTSPTLYDCHS 540
 QY 541 MKGNQIMTKYRKDXTLVHPVSGSCMDCSESDHRIFMNTCPSSLTQOMLFEHTNSTVLEKF 600
 DB 541 MKGNQIMTKYRKDXTLVHPVSGSCMDCSESDHRIFMNTCPSSLTQOMLFEHTNSTVLEKF 600
 QY 601 NRN 603
 DB 601 NRN 603

RESULT 3
 BAD21405 PRELIMINARY; PRT; 634 AA.
 ID BAD21405
 AC BAD21405
 DT 01-JUN-2004 (T-EMBLrel. 27, Created)
 DT 01-JUN-2004 (T-EMBLrel. 27, Last sequence update)
 DE MFLJ0205 protein (Fragment).
 GN MFLJ0205.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBTAXID=10090;
 RN NCBTAXID=10090;
 RP SEQUENCE FROM N.A.
 RA Okazaki N., Kitano R., Ohara R., Inamoto S., Koseki H., Hirakawa S.,
 RA Saga Y., Kitamura H., Nakagawa T., Nagase T., Ohara O., Koga H.,
 RT "Prediction of the Coding Sequences of 110 Mouse Homologues of FlJ Gene:
 RT The Complete Nucleotide Sequences of 110 Mouse FlJ-homologous cDNAs
 RT Identified by Screening of Terminal sequences of cDNA Clones Randomly
 RT Sampled from size-fractionated libraries";
 RT Submitted (FEJ-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL/AL313155; BAD21405.1; -.
 FT NON_TER 1

SQ SEQUENCE 634 AA; 72315 MW; C79B82D4D0052C81 CRC64;
 Query Match 95.7%; Score 3138; DB 2; Length 634;
 Best Local Similarity 95.5%; Pred. No. 7.6e-246;
 Matches 576; Conservative 9; Mismatches 18; Indels 0; Gaps 0;

QY 1 MRKEKRLLOAVLVAALVLLPNVGLMALYERQPDGTGGGAAPAAAGGSHSRK 60
 DB 32 MRKEKRLLOAVLVAALVLLPNVGLMALYERQPDGTGGGAAPAAAGGSHSRK 91
 QY 61 KTFPLGDGQKLDKMDHKEAIRDAQVNGEGQRPVPMTDABRVDAVRENGFNIVSDK 120
 DB 61 KTFPLGAEQRLKDMHKEAIRDAQVNGEGQRPVPMTDABRVDAVRENGFNIVSDK 151
 QY 92 KTFPLGAEQRLKDMHKEAIRDAQVNGEGQRPVPMTDABRVDAVRENGFNIVSDK 151
 DB 92 KTFPLGAEQRLKDMHKEAIRDAQVNGEGQRPVPMTDABRVDAVRENGFNIVSDK 151
 QY 121 ISLNRSLPDIRHPNCNKRYLETLPNTSIIIPHNEGSSLLRTVSVLNRSPELVAEI 180
 DB 152 ISLNRSLPDIRHPNCNKRYLETLPNTSIIIPHNEGSSLLRTVSVLNRSPELVAEI 211
 QY 181 VLVDDSDREHLKKPLEDYMALPPSVRIILRTKREGILRTMLGASAAATGDTITFLDSHC 240
 DB 212 VLVDDSDREHLKKPLEDYMALPPSVRIILRTKREGILRTMLGASAAATGDTITFLDSHC 271
 QY 241 EAVNMVLPPLDLRIANRRTIVCPMIDVIDHDPRYETQAGDAMRGAFPMEMYKRIPI 300
 DB 272 EAVNMVLPPLDLRIANRRTIVCPMIDVIDHDPRYETQAGDAMRGAFPMEMYKRIPI 331
 QY 301 PELQKADPSDPESPVMAGGLFVADRKKFWELGSDYDGLGIEWGEGYEISFKVMCGGRM 360
 DB 332 PELQKADPSDPESPVMAGGLFVADRKKFWELGSDYDGLGIEWGEGYEISFKVMCGGRM 391
 QY 361 EDIPCSRVGHIYRKYPYVPAGVSLARNLKRVAEVMDEYAEYIQRREYRHLSAGDV 420
 DB 392 EDIPCSRVGHIYRKYPYVPAGVSLARNLKRVAEVMDEYAEYIQRREYRHLSAGDV 451
 QY 421 AVQKRLRSSLNCKSFPMFTKIAMDLPKFYPPVEPPAAWGIIRNVTGLCADTKGALG 480
 DB 452 AVQKRLRSSLNCKSFPMFTKIAMDLPKFYPPVEPPAAWGIIRNVTGLCADTKGALG 511
 QY 481 SPRLLETCIRGGEAAMNMQVFTFTWRREDIRPGDPQHTKKCFDAVSHTSPTLYDCHS 540
 DB 512 SPRLLETCIRGGEAAMNMQVFTFTWRREDIRPGDPQHTKKCFDAVSHTSPTLYDCHS 571
 QY 541 MKGNQIMTKYRKDXTLVHPVSGSCMDCSESDHRIFMNTCPSSLTQOMLFEHTNSTVLEKF 600
 DB 572 MKGNQIMTKYRKDXTLVHPVSGSCMDCSESDHRIFMNTCPSSLTQOMLFEHTNSTVLEKF 631
 QY 601 NRN 603
 DB 632 NRN 634

RESULT 4
 GL10_MOUSE STANDARD; PRT; 603 AA.
 ID GL10_MOUSE
 AC Q6P9S7; Q6RAQ2; Q8B208; Q91YJ6;
 DT 01-OCT-2004 (Rel. 45, Created)
 DT 01-OCT-2004 (Rel. 45, Last sequence update)
 DE 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Polypeptide N-acetylglucosaminyltransferase 10 (EC 2.4.1.41)
 DE (Protein-UDP acetylglucosaminyltransferase 10 (UDP-
 DE GAlNAc:Polypeptide N-acetylglucosaminyltransferase 10) (Polypeptide
 DE GAlNAc transferase 10) (GAlNAc-T10) (pp-GAlNAc 10).
 GN Name=GAlNAcT10;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBTAXID=10090;
 RN NCBTAXID=10090;
 RP SEQUENCE FROM N.A.
 RA Okazaki N., Kitano R., Ohara R., Inamoto S., Koseki H., Hirakawa S.,
 RA Saga Y., Kitamura H., Nakagawa T., Nagase T., Ohara O., Koga H.,
 RT "Prediction of the coding sequences of 110 mouse FlJ-homologous cDNAs
 RT the complete nucleotide sequences of 110 mouse FlJ-homologous cDNAs

RT identified by screening of terminal sequences of cDNA clones randomly
 RT sampled from size-fractionated libraries.",
 RL DNA Res. 11:167-180(2004).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6, and FVB/N; TISSUE=Brain, and Breast tumor;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loughellano N.A., Peters G.J., Abramson R.D., Mulhaly S.J.,
 RA Bosak S.A., McEwen P.D., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richardson S., Wooley K.C., Hale S., Garcia A.M., Gay L.J., Hultky S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahney J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,
 RA Butterfield J.S.N., Krzywinski M.I., Skalek A., Smalins D.E.,
 RA Schneroch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.",
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [3]
 RN SEQUENCE OF 76-603 FROM N.A.
 RP STRAIN=C57BL/6J; TISSUE=Colon;
 RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 RA Nikaido I., Osato N., Saito R., Suzuki H., Yananaka I., Kiyosawa H.,
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojohori T.,
 RA Balderelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 RA Shiraki L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
 RA Blake J.A., Bradt D., Brusic V., Chochia C., Corbani L.E., Cousins S.,
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
 RA Gaasterland T., Gariboldi M., Glaser C., Godzik A., Gough J.,
 RA Grønborg S., Guenlinclon S., Hirokawa N., Jackson I.J., Jarvis E.D.,
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
 RA Nagashima T., Nimata K., Okido T., Pavan W.J., Pettes G., Pesole G.,
 RA Petrovsky N., Pillai R., Pontius U.U., Qi D., Ramachandran S.,
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
 RA Sanderlin A., Schneider C., Sempile C.A., Setou M., Shimada K.,
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
 RA Varrault R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 RA Wilming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
 RA Hironaka-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,
 RA Shiraki T., Waki K., Kawai J., Atzawa K., Arakawa T., Fukuda S.,
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
 RA Yamanishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
 RA Birney E., Hayashizaki Y.,
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs",
 RL Nature 420:563-573(2002).
 [4]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=21264503; PubMed=11278534; DOI=10.1074/jbc.M009638200;
 RA Ten Hagen K.G., Bedi G.S., Tetsch D., Kingsley P.D., Hagen F.K.,
 RA Balys M.M., Beres T.M., Degand P., Tabak L.A.,
 RT "Cloning and characterization of a ninth member of the UDP-
 RT GalNAc:polypeptide N-acetylglucosaminyltransferase family,
 RT ppgantase-T9",
 RL J. Biol. Chem. 276:17395-17404(2001).
 CC -1- FUNCTION: Catalyzes the initial reaction in O-linked
 CC oligosaccharide biosynthesis, the transfer of an N-acetyl-D-
 CC galactosamine residue to a serine or threonine residue on the
 CC protein precursor. Has activity toward Muc5Ac and EA2 peptide

CC substrates (By similarity).
 CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =
 CC UDP + N-acetyl-D-galactosaminyl-polypeptide.
 CC -1- COFACTOR: Manganese and calcium (By similarity).
 CC -1- PATHWAY: Glycosylation.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By
 CC similarity).
 CC -1- TISSUE SPECIFICITY: Expressed at higher level than GALNT9. In the
 CC developing hindbrain region of E14.5 embryos it accumulates in the
 CC rapidly dividing, undifferentiated ventricular zone adjacent to
 CC the pons. It also accumulates in the regions immediately rostral
 CC and caudal to the dorsal rhombic lips differentiating into the
 CC cerebellum. Not expressed in the developing choroid plexus.
 CC -1- DOMAIN: There are two conserved domains in the glycosyltransferase
 CC region: the N-terminal domain (domain A, also called GT1 motif),
 CC which is probably involved in manganese coordination and substrate
 CC binding and the C-terminal domain (domain B, also called
 CC Gal/GalNAc-T motif), which is probably involved in catalytic
 CC reaction and UDP-gal binding (By similarity).
 CC -1- DOMAIN: The ricin B-type lectin domain binds to GalNAc and
 CC contributes to the glycopeptide specificity (By similarity).
 CC -1- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T
 CC subfamily.
 CC -1- SIMILARITY: Contains 1 ricin B-type lectin domain.
 CC -1- CAUTION: According to experiments made in rat, this enzyme is
 CC unable to transfer GalNAc onto serine or threonine residue on the
 CC protein receptor, but instead requires the prior addition of a
 CC GalNAc on a peptide before adding additional GalNAc moieties,
 CC thereby acting as a glycopeptide transferase.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, AK131155; BA021405.1; ALT_INIT.
 CC EMBL, BC016585; AAH16585.1; ALT_INIT.
 CC DR EMBL, BC060617; AAH0617.1; -.
 CC DR EMBL, AK033515; BA028334.1; -.
 CC DR HSSP, P26514; 1KNM.
 CC DR MGD, MGI:1890480; Galnt10.
 CC DR GO, GO:0004653; F:polypeptide N-acetylgalactosaminyltransferase. . . ; IDA.
 CC GO, GO:0006493; P:O-linked glycosylation; IDA.
 CC DR InterPro, IPR001173; Glyco trans 2.
 CC DR InterPro, IPR008997; RicinB like.
 CC DR InterPro, IPR000772; Ricin B lectin.
 CC DR Pfam, PF00535; Glycos transf_2; 1.
 CC DR Pfam, PF00652; Ricin B lectin; 2.
 CC DR SMART, SM00458; RICIN_1.
 CC DR PROSITE, PSS0231; RICIN B LECTIN; 1.
 CC DR Calcium; Glycosyltransferase; Golgi stack; Lectin; Manganese;
 CC KEGG, K01101; Glycosyltransferase; Transmembrane.
 CC FT DOMAIN 1 11 Cytoplasmic (Potential).
 CC FT TRANSMEM 12 31 Signal-anchor for type II membrane
 CC FT protein (Potential).
 CC FT DOMAIN 144 253 Lumenal (Potential).
 CC FT DOMAIN 311 373 Catalytic subdomain A.
 CC FT DOMAIN 458 590 Ricin B-type lectin.
 CC FT DISULFID 471 488 By similarity.
 CC FT DISULFID 523 538 By similarity.
 CC FT CARBOHYD 124 124 N-linked (GlcNAc . . .) (Potential).
 CC FT CARBOHYD 146 146 N-linked (GlcNAc . . .) (Potential).
 CC FT CARBOHYD 593 593 N-linked (GlcNAc . . .) (Potential).
 CC FT CONFLICT 233 233 V -> I (in Ref. 1).
 CC SQ SEQUENCE 603 AA; 6916 MW; FF55FBAVEID7544 CRC64;
 CC Query Match 95.7%; Score 3137; DB 1; Length 603;
 CC Best Local Similarity 95.4%; Pred. No. 8.6e-246;

Matches 575; Conservative 10; Mismatches 18; Indels 0; Gaps 0;

QY 1 MRKEKRLQAVLVAALVLPVNGLMALYREPODPTGCGSAAVAPAAAGGSHSROK 60
 DB 1 MRKEKRLQAVLVAALVLPVNGLMALYREPODPTGCGSAAVAPAAAGGSHSROK 60
 QY 61 KTFELGDDGQKLDKMDHKEAIRDAQVNGEGQRPYPMTDAERVDQAYRENGENIYVSDK 120
 DB 61 KTFELGDDGQKLDKMDHKEAIRDAQVNGEGQRPYPMTDAERVDQAYRENGENIYVSDK 120
 QY 121 ISLNRLPDIRHNCNCKRYLETLPNTSIIIFPHNEGSSLLRTYHVSVNRSPPELVARI 180
 DB 121 ISLNRLPDIRHNCNCKRYLETLPNTSIIIFPHNEGSSLLRTYHVSVNRSPPELVARI 180
 QY 181 VLVDPSDREHKKPLEDYMALPFSYRIARTKKREGILRTMKGASVATGDTTFPLDSHC 240
 DB 181 VLVDPSDREHKKPLEDYMALPFSYRIARTKKREGILRTMKGASVATGDTTFPLDSHC 240
 QY 241 EAVNWMPLPLDLRIANRRTIVCPMIDVIDHDDFRYETQAGDAMRGAFDWMYYKRIPIR 300
 DB 241 EAVNWMPLPLDLRIANRRTIVCPMIDVIDHDDFRYETQAGDAMRGAFDWMYYKRIPIR 300
 QY 301 PELQKADPSDPSPSPVMAAGLFAVDRKMFELGYPGLEINGEGYEISFKVMCGGRM 360
 DB 301 PELQKADPSDPSPSPVMAAGLFAVDRKMFELGYPGLEINGEGYEISFKVMCGGRM 360
 QY 361 EDIPCSRVGHIYRKYPYKVPAGVSLARNLKRVAEVMDEYAEIYQRRPEYRHLISAGDV 420
 DB 361 EDIPCSRVGHIYRKYPYKVPAGVSLARNLKRVAEVMDEYAEIYQRRPEYRHLISAGDV 420
 QY 421 AVOKLRSSLNCKSPKFMFTKIAMDLPKFYPPVEPAAAGSIRNVGTGLCADTKRGALG 480
 DB 421 AVOKLRSSLNCKSPKFMFTKIAMDLPKFYPPVEPAAAGSIRNVGTGLCADTKRGALG 480
 QY 481 SPLRLGCGVRGGEAAAMNNQVFTFTWRREDIRPGDQHTKCFPDASISHTSPITYDCIS 540
 DB 481 SPLRLGCGVRGGEAAAMNNQVFTFTWRREDIRPGDQHTKCFPDASISHTSPITYDCIS 540
 QY 541 MKGNQLMKRYKDKTLVHPVSGSCMDCSESDHRIFMNTCNPSLTOOMLFEHTNSIYLEK 600
 DB 541 MKGNQLMKRYKDKTLVHPVSGSCMDCSESDHRIFMNTCNPSLTOOMLFEHTNSIYLEK 600
 QY 601 NRN 603
 DB 601 NRN 603

RESULT 5
 AAH60617 PRELIMINARY; PRT; 603 AA.

AC AAH60617
 DT 12-MAY-2004 (TEMBLrel. 27, Created)
 DT 12-MAY-2004 (TEMBLrel. 27, Last sequence update)
 DE UDP-N-acetyl-alpha-D-galactosamine:polypeptide
 DE N-acetyl-galactosaminyltransferase 10.
 GN GALNT10.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBT_TaxID=10090;
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Brain;
 MEDLINE=22386257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altshuler S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haileh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ueda T.B., Tomihyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richard S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Vailion D.K., Murry D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettlemen M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting J., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Gilmwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzyzanski M.I., Skalska U., Smalins D.E., Scherch A., Schein J.E.,
 RA Jones S.J., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RA Strausberg R.;
 RL Submitted (Oct-2003) to the EMBL/GenBank/DBJ database.
 DR EMBL, BC060617, AAH60617.1, -.
 KW transferase.
 SQ SEQUENCE 603 AA; 6916 MW; FF55FBA7E1DD7544 CRC64;

Query Match 95.7%; Score 3137; DB 2; Length 603;
 Best Local Similarity 95.4%; Pred. No. 8,6e-246;
 Matches 575; Conservative 10; Mismatches 18; Indels 0; Gaps 0;

QY 1 MRKEKRLQAVLVAALVLPVNGLMALYREPODPTGCGSAAVAPAAAGGSHSROK 60
 DB 1 MRKEKRLQAVLVAALVLPVNGLMALYREPODPTGCGSAAVAPAAAGGSHSROK 60
 QY 61 KTFELGDDGQKLDKMDHKEAIRDAQVNGEGQRPYPMTDAERVDQAYRENGENIYVSDK 120
 DB 61 KTFELGDDGQKLDKMDHKEAIRDAQVNGEGQRPYPMTDAERVDQAYRENGENIYVSDK 120
 QY 121 ISLNRLPDIRHNCNCKRYLETLPNTSIIIFPHNEGSSLLRTYHVSVNRSPPELVARI 180
 DB 121 ISLNRLPDIRHNCNCKRYLETLPNTSIIIFPHNEGSSLLRTYHVSVNRSPPELVARI 180
 QY 181 VLVDPSDREHKKPLEDYMALPFSYRIARTKKREGILRTMKGASVATGDTTFPLDSHC 240
 DB 181 VLVDPSDREHKKPLEDYMALPFSYRIARTKKREGILRTMKGASVATGDTTFPLDSHC 240
 QY 241 EAVNWMPLPLDLRIANRRTIVCPMIDVIDHDDFRYETQAGDAMRGAFDWMYYKRIPIR 300
 DB 241 EAVNWMPLPLDLRIANRRTIVCPMIDVIDHDDFRYETQAGDAMRGAFDWMYYKRIPIR 300
 QY 301 PELQKADPSDPSPSPVMAAGLFAVDRKMFELGYPGLEINGEGYEISFKVMCGGRM 360
 DB 301 PELQKADPSDPSPSPVMAAGLFAVDRKMFELGYPGLEINGEGYEISFKVMCGGRM 360
 QY 361 EDIPCSRVGHIYRKYPYKVPAGVSLARNLKRVAEVMDEYAEIYQRRPEYRHLISAGDV 420
 DB 361 EDIPCSRVGHIYRKYPYKVPAGVSLARNLKRVAEVMDEYAEIYQRRPEYRHLISAGDV 420
 QY 421 AVOKLRSSLNCKSPKFMFTKIAMDLPKFYPPVEPAAAGSIRNVGTGLCADTKRGALG 480
 DB 421 AVOKLRSSLNCKSPKFMFTKIAMDLPKFYPPVEPAAAGSIRNVGTGLCADTKRGALG 480
 QY 481 SPLRLGCGVRGGEAAAMNNQVFTFTWRREDIRPGDQHTKCFPDASISHTSPITYDCIS 540
 DB 481 SPLRLGCGVRGGEAAAMNNQVFTFTWRREDIRPGDQHTKCFPDASISHTSPITYDCIS 540
 QY 541 MKGNQLMKRYKDKTLVHPVSGSCMDCSESDHRIFMNTCNPSLTOOMLFEHTNSIYLEK 600
 DB 541 MKGNQLMKRYKDKTLVHPVSGSCMDCSESDHRIFMNTCNPSLTOOMLFEHTNSIYLEK 600
 QY 601 NRN 603
 DB 601 NRN 603

RESULT 6
 Q700E9 PRELIMINARY; PRT; 644 AA.

FT DISULFID 532 549 By similarity.
 FT DISULFID 575 591 By similarity.
 FT CARBOHYD 136 136 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 170 170 N-linked (GlcNAc...) (Potential).
 SQ SEQUENCE 622 AA; 71811 MW; F8A3A21159C7B52 CRC64;

Query Match 42.5%; Score 1392.5; DB 1; Length 622;
 Best Local Similarity 48.8%; Pred. No. 4.5e-104;
 Matches 279; Conservative 87; Mismatches 175; Indels 31; Gaps 14;

QY PAAGQSHSRQKTFPLGCGQKLDKMDKAIKRDQAVNGQGRPYPTMDERDQ--106
 DB PPALDDEALDPEKRYGRH-EKIK-WEDEAAYEKERKREBPGRVVKLPEDKEVEKA 116
 QY 107 --AYRENGENIYVSDKISLNRSLPDRIHPNCKSKRYLETLPNTSIILPFHNEGMSILLRT 164
 DB 117 LSIYKANGNAYISDMISLNRSLKIDIRHCKCKMMNTSAKLPTYSVIFPHIEHNSILLNS 176
 QY 165 VHSVLNRSPPELVAEIVLVDFSDRHLKKPLEDYM--ALPFSVILATKKREGILRT 221
 DB 177 VYSVNRSPPELKEIILVDFSEKDALRQLEDPLKKNKIDHIVKVLRTKKEGILRGR 236
 QY 222 MGLASVATGDTVITFLSHGCAANNWMLPRLDRIARNRKTIYCVGMIDVIDHDDPRYETQAG 281
 DB 237 QLGADATGEIILFLDASHANNNWMLPRLDPIAEYRTVCPVVIDCETYEVPQ-D 295
 QY 282 DAMRGAFDMEYKRIPIPELOKADPSDFESPVWAGLFAVDKRMFELGGYDGLRI 341
 DB 296 EGRGSGFDNAFYKRIPLTKK-DRESPTKRFNSPVHGGYFALSAKMFELGGYDGLDI 354
 QY 342 WGGQYELSPKVMCGRMEDIPCSRHYIYR-KYVPYK-VPAVSLARLTKFAEVMND 399
 DB 355 WGGQYELSPKVMCGRMEDIPCSRHYIYR-KYVPYK-VPAVSLARLTKFAEVMND 414
 QY 400 EYAEIYQRRPERRHLSAGVANOKLRSLNCKSKRMFKTKAMLPKRYPPVEPPAA 459
 DB 415 DYKRETIYKRRPGVGNADAGLKLKMKIIRKLOCKSPDFWKEIAPFODKYPVEPPKA 474
 QY 460 WGEIRVGTGLCADTKHGAISPLREGCVR--GRGANNNMNOVFETFMEDIRPG 515
 DB 475 EGEIRVGTGLCADTKHGAISPLREGCVR--GRGANNNMNOVFETFMEDIRPG 515
 QY 516 PHTKKFCFDALISHT--SPVTLVDCHSMKGNQMLKTR-KDKTLVHPVSGSCMDSCSDHR 572
 DB 527 --KGRKICFDSCSTSVKAPVILFDCHSMKGNQMLFKTRVAKQIYHPIISGGLVADENK 584
 QY 573 -IFMNTCNPSLQQLFHEHTNSTVLEKFRN 603
 DB 585 FLHMKKCDSSDLQKMAQTVDNELLETRQAN 616

RESULT 8
 GLT6_DROME STANDARD; PRT; 666 AA.
 AC Q6WV16; Q9SR40; Q9VZK5;
 DT 01-OCT-2004 (Rel. 45, last sequence update)
 DT 01-OCT-2004 (Rel. 45, last annotation update)
 DE N-acetyl-galactosaminyltransferase 6 (EC 2.4.1.-) (Protein-UDP
 acetyl-galactosaminyltransferase 6) (UDP-GalNAc:polypeptide N-
 acetyl-galactosaminyltransferase 6) (pp-GANTase 6).
 GN NameSpace6; ORFNames=CG2103;
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OC NCB1_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A., ENZYME ACTIVITY, TISSUE SPECIFICITY, AND
 RP DEVELOPMENTAL STAGE.
 RC STRAIN=Canton-S; TISSUE=Embryo;
 RX MEDLINE=22841110; PubMed=12829714; DOI=10.1074/jbc.M303836200;
 TX Ten Hagen K.G., Tran D.T., Gerken T.A., Stein D.S., Zhang Z.,

RT "Functional characterization and expression analysis of members of the
 RT UDP-GalNAc:polypeptide N-acetyl-galactosaminyltransferase family from
 RT Drosophila melanogaster.";
 RL J. Biol. Chem. 278:35039-35048(2003).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
 RA Adams M.D., Cealinker S.B., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.B., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abrial J.F., Abgaryan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Curry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostler C., Gabrielian A.E., Garg N.S., Galibert W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey K.A., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kenton J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasok P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Martel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Stykars R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wesserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RN Science 287:2185-2195(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley; TISSUE=Embryo;
 RX MEDLINE=22426066; PubMed=12537569;
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
 RA George R.A., Guarnier H., Krommiller B., Pacle J.M., Park S., Wan K.H.,
 RA Rubin G.M., Cealinker S.B.;
 RT "A Drosophila full-length cDNA resource.";
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
 CC -1- FUNCTION: Glycopeptide transferase involved in O-linked
 CC oligosaccharide biosynthesis, which catalyzes the transfer of an
 CC N-acetyl-D-galactosamine residue to an already glycosylated
 CC peptide. In contrast to other proteins of the family, it does not
 CC act as a peptide transferase that transfers GalNAc onto serine or
 CC threonine residue on the protein receptor, but instead requires
 CC the prior addition of a GalNAc on a peptide before adding
 CC additional GalNAc moieties. Some peptide transferase activity is
 CC however not excluded, considering that its appropriate peptide
 CC substrate may remain unidentified. Prefers the diglycosylated
 CC Nucleo-3/13 as substrate.
 CC -1- Cofactor: Manganese and calcium (By similarity).
 CC -1- PATHWAY: Glycosylation.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By
 CC similarity).
 CC -1- TISSUE SPECIFICITY: In embryo, it is specifically expressed in
 CC the salivary glands from stage 12, becoming stronger at stage 13.
 CC Not expressed in other tissues. Further expressed during

```

CC oogenesis, in the somatically derived follicle cells that surround
CC the developing oocyte, which are involved in the maturation of the
CC oocyte and construction of the egg shell, as well as playing a
CC role in subsequent embryonic pattern formation.
CC - DEVELOPMENTAL STAGE: Expressed throughout embryonic, larval, pupal
CC and adult stages, with increasing levels during larval
CC development.
CC - DOMAIN: There are two conserved domains in the glycosyltransferase
CC region: the N-terminal domain (domain A, also called GT1 motif),
CC which is probably involved in manganese coordination and substrate
CC binding and the C-terminal domain (domain B, also called
CC Gal/GalNAc-T motif), which is probably involved in catalytic
CC reaction and UDP-Gal binding (By similarity).
CC - DOMAIN: The ricin B-type lectin domain binds to GalNAc and
CC contributes to the ricin B-type lectin domain specificity (By similarity).
CC - SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T
CC subfamily.
CC - SIMILARITY: Contains 1 ricin B-type lectin domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
CC DR EMBL; AY268067; AAC06703.1; -
CC DR EMBL; AE003476; AAF47690.1; -
CC DR EMBL; AY061629; AAL29177.1; -
CC DR FlyBase; FBgn0035375; pgant6.
CC DR InterPro; IPR001173; Glyco_transf_2.
CC DR InterPro; IPR008997; RicinB_1like.
CC DR InterPro; IPR00772; RicinB_1lectn.
CC DR Pfam; PF00535; Glycos_transf_2_1.
CC DR Pfam; PF00652; RicinB_1lectin; 3.
CC DR SMART; SM00458; RICIN; 1.
CC DR PROSITE; PS00331; RICIN_B_HECTIN; 1.
CC DR KEGG; Glycosyltransferase; Golgi stack; Lectin; Manganese;
CC KW Signal-anchor; Transferase; Transmembrane.
CC FT DOMAIN 1 11
CC FT TRANSMEM 12 31
CC FT Signal-anchor for type II membrane
CC FT protein (Potential).
CC FT DOMAIN 32 666
CC FT DOMAIN 201 311
CC FT DOMAIN 367 429
CC FT DOMAIN 518 648
CC FT DISULFID 531 548
CC FT DISULFID 577 594
CC FT DISULFID 621 636
CC FT CARBOHYD 181 181
CC FT CARBOHYD 285 285
CC FT CARBOHYD 651 651
CC FT CARBOHYD 657 657
CC FT CONFLICT 95 96
CC FT CONFLICT 107 107
CC FT CONFLICT 499 499
CC FT CONFLICT 666 AA; 76972 MW; CACGACGEA860600C CRR64;
CC SEQUENCE
CC
CC Query Match 41.5%; Score 1360; DB 1; Length 666;
CC Best Local Similarity 49.0%; Pred. No. 2,2e-101;
CC Matches 277; Conservative 76; Mismatches 160; Indels 52; Gaps 15;
CC
CC 67 DGQRLKMDHDKRAIRDAORVNGSGQRPYPMTDAERDQAVR---ENGFNIVYSDKSL 123
CC DB DASVKKMDHDIYFMKDKARVGLGCGKASTLDDSGQRLKRGMSLENGFNALSDSISV 180
CC 124 NRSLEPIRHPCNCSKRYLETLPNTSIIIPFHNEGSSLLRTVSHVYLNRSPELVAEIVLV 183
CC DB NRSVPDIHPIPCRKKEVYAKLPTVSVIIIFNVEYISVLMSRSHSLNNSPELMKEIILV 240
CC 184 DDFSREHLKKPLEDIYMA-LPFSVRIILRTKKEGLIRRMUGASVATGDVITFDLSHCEA 242

```

Dd		241	DHSDREYLGKELTYIAEHFKWVRVRLPRTGLIGARAAGANNAVAEVLIFLDSHEA	300
Qy		243	NVMNLPPLDLRIANNRKTIVCEMTDIVDDHDFFRYETQAADMARCAFDEMYTKIPIPE	302
Dd		301	NYNNLPPLLEBIALNKRTAVCPFIDVIDHTFHRAQ-DEGARGAFMEFPFKPLPLE	359
Qy		303	LQXKDSPSPESPWAGSLFAVDKKFMELGGYDPSGLEIWCGEGYEISPKWMCGMRNE	362
Dd		360	DLK-HPADPFPSPIWAGGLEAISSEFPWEJGGYDEGLDIWGGEYELSFKIMCGGEYD	418
Qy		363	IPCSRVGHITV-----KYVPKYPAVGSILANLRKVAEVMMDAEAYEITYORPE-YRHLS	416
Dd		419	APCRIGHITVGRPNHQSPRK---GDYLHNTRYKVAEVMMDDEVKNLYLSHGDDLYESVD	475
Qy		417	AGDAVAVQKLRSINCKSPKFWMTKIAMDLPKFTPYVEPPPAANGIRNVGT-GLCADT-	474
Dd		476	PGLDTLEGKAITRTKNCSPKFMEFEVAFDMLKTYPPDPSPSYAMGALQNWNQNLCDTL	535
Qy		475	---FHGALG-----SPIRLGECVRGRGEAAANNMQVTFTFRARDIRPGDQHKKF	522
Dd		536	GRRKHNRKMGYACADNIKTPOK-----TFQEWLSWKEDLR---LRRKGE	576
Qy		523	CFDA--ISHTSPVLYCYCHSKMGQLMKY-RKDXTLYHPVSG-SCMCSESDEHRIFMNTC	578
Dd		577	CLDVQIDYDANA.PVLMWDCHSGGNGQYYTYDRHKQLGHTEGRRCLELLPSPQEVANKC	636
Qy		579	NPSSLTOQMLEPHTNSTVLEKEFNEN	603
Dd		637	: : : : : DTDNRFQOOWNFGSFNKTAIDWYSQD	661
RESULT 9				
AAQS6703				
ID	AAQS6703	PRELIMINARY;	PRT;	666 AA.
AC	AAQS6703:			
DT	02-MAR-2004 (TREMBLrel. 27, Created)			
DT	02-MAR-2004 (TREMBLrel. 27, Last sequence update)			
DT	02-MAR-2004 (TREMBLrel. 27, Last annotation.update)			
DE	UDP-GalNAc:polypeptide N-acetylglactosaminyltransferase (EC			
DE	2.4.1.41).			
GN	PGANT6.			
OS	Drosophila melanogaster (fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	PubMed:12829714;			
RA	Ten Hagen K.G., Tran D.T., Gerken T.A., Stein D.S., Zhang Z.;			
RA	"Functional Characterization and Expression Analysis of Members of the			
RT	UDP-GalNAc:Polypeptide N-Acetylglactosaminyltransferase Family from			
RL	J. Biol. Chem. 278:35039-35048 (2003).			
DR	EMBL; AY268067; AAQ6703.1;			
KM	Glycoyltransferase; Transferase.			
SQ	SEQUENCE 666 AA; 76958 MW; ID18362EE0DBD196 CRC64;			
Query Match	41.5%; Score 1360; DB 2; Length 666;			
Best Local Similarity	49.0%; Pred. No. 2.2e-101;			
Matches 277; Conservative	76; Mismatches 160; Indels 52; Gaps 15;			
Qy		67	DGOKLKMDHDEKALRRDAQRVVGNGEOGRPRYPMTDAERYDAUR--ENGFIYVSDKSL	123
Dd		121	DASYVKDMNDTTPEKDKAKVGLGEGGKASTLDDESQRDLKRMSTLNGRNALLSDSISV	180
Qy		124	NRSLDPDIRHPNCNSKRLEYLTLPNTSIIPFHNEGSSILRTVHSVILNRSPELVAAIVLV	183
Dd		181	NRSPVDIRHPLCRKKEKVAKLPVTSVIIIFYNEYLVTRMSVSHLIRSPELMKEIIIV	240
Qy		184	DDFSDEHLKRPLEDYMA-LPFSRIILRTKKRGSELIRTMGLASVAVGVITFLDSHCA	242
Dd		241	DHSDREYLGKELTYIAEHFKWVRVRLPRTGLIGARAAGANNAVAEVLIFLDSHEA	300

QY 189 REHLKKPELDYML-PESVRLRTKKEGLIRTRMLGASVATGDPVITFLDSHCANVNL 247
 DB 147 KEFLHNLEEDYKONLPKVLAVROPERTGLIKARLAQAKIASGVLFLDSHTAGVNL 206
 QY 248 PPLIDRIARRKRTIVCEMIDVIDIDHDFRYETQAGDANRGAFDWEMYKRIPIPELOKAD 307
 DB 207 PPLLEPIAENPKTCVCPILIDVIDDQTFDVAHQ-DEGGRLFDWTFHYKRVIKNE-DRIS 264
 QY 308 PSDPESPPVMAAGGFAVADRKKFMELGSGYDELITWGEQVEISKVMCGGRMEDICSR 367
 DB 265 PTEFPSPVMAAGGFAIGADFPWELGSGYDELITWGEQVEISKVMCGGRMEDICSR 324
 QY 368 VGHIRKVPYKPVAGVS-LARNIKRVAEVMDEYATVIOBRE-RRHLSAGDVAQKK 425
 DB 325 FGHIRYTPSPSPSRKDTFTRNHKKRVAETIMDEYKQYITDRDERAKTADAGMSCKKT 384
 QY 426 LRSLNCKSEKFMFTKIAMDLPKFYVPEPPAAAGBIRNVG-TGLCADTKGALGSPLR 484
 DB 385 IREKLMCKPKFMFLQEVAPRIELIYPPVEPEPPVASSIGVADSSLCIDTMQRGRGPIG 444
 QY 485 LECGVRGRGAANNVQVFTFTWEDIRPGDPQTKP---CPDAISHT--SPVTLVYDCH 539
 DB 445 LYPGNSNLIIEPT-NHNOYFVHSMHRDI-----OH--KYGECFPDVPQSKDGPVTITFTCH 496
 QY 540 SMKGNOLMKYRKOKTLVHPVSGCMD-----CSSEDRIFMNTCNPSLTOQWLFEBHNTS 595
 DB 497 MHQGNQFOY-DHKIQQIKRNGVCTIDSDPNAKECHRNYSKQLPQTSITVPPFDEHWST 555
 QY 596 VL 597
 DB 556 LL 557

RESULT 12

GLT4 DROME STANDARD; PRT; 659 AA.

AC 081A42; 081011;
 DT 01-OCT-2004 (Rel. 45, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE N-acetyl-galactosaminyltransferase 4 (EC 2.4.1.-) (Protein-UDP
 DE acetyl-galactosaminyltransferase 4) (UDP-GalNAc:polypeptide N-
 DE acetyl-galactosaminyltransferase 4) (pp-GalNAc 4).
 CN Name:pgant4; ORFNames:CG31956;
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 [1]
 RN SEQUENCE FROM N.A., ENZYME ACTIVITY, TISSUE SPECIFICITY, AND
 RP DEVELOPMENTAL STAGE.
 RC STRAIN=Canton-S; TISSUE=Embryo;
 RX MEDLINE=28841110; PubMed=12829714; DOI=10.1074/jbc.M3036200;
 RA Ten Hagen K.G., Tran T.T., Gecken T.A., Stein D.S., Zhang Z.;
 RT "Functional characterization and expression analysis of members of the
 RT UDP-GalNAc:polypeptide N-acetyl-galactosaminyltransferase family from
 RT Drosophila melanogaster."
 RL J. Biol. Chem. 278:35039-35048 (2003).
 [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.B., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abiri J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Baau A., Baxendale J., Bayraktaroglu L., Beasley S.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Cherry K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Goadec A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiland T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegam C.,
 RA Isailin M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Koditz C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laske P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Morkulov G., Mshelina N.V., Mobery C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirbax R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195 (2000).
 [3]
 RN REVISIONS.
 RP MEDLINE=2426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.B.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Beutenkourt B.R., Celisner S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 RT systematic review."
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22 (2002).
 [4]
 RN SEQUENCE OF 16-659 FROM N.A.
 RX PubMed=11925450; DOI=10.1074/jbc.M20264200;
 RA Schwientek T., Bennett E.P., Flores C., Thacker J., Hollmann M.,
 RA Reis C.A., Behrens J., Mandel U., Kech B., Schaefer M.A.,
 RA Haselmann K., Zubarev R., Roepstorff P., Burchell J.M.,
 RA Taylor-Papadimitriou J., Hollingsworth M.A., Clausen H.;
 RT "Functional conservation of subfamilies of putative UDP-N-
 RT acetyl-galactosamine:polypeptide N-acetyl-galactosaminyltransferases in
 RT Drosophila, Caenorhabditis elegans, and mammals. One subfamily
 RT composed of 1(2)35aa is essential in Drosophila."
 RL J. Biol. Chem. 277:22623-22638 (2002).
 CC -1- FUNCTION: Glycopeptide transferase involved in O-linked
 CC oligosaccharide biosynthesis, which catalyzes the transfer of an
 CC N-acetyl-D-galactosamine residue to an already glycosylated
 CC peptide. In contrast to other proteins of the family, it does not
 CC act as a peptide transferase that transfers GalNAc onto serine or
 CC threonine residue on the protein receptor, but instead requires
 CC the prior addition of a GalNAc on a peptide before adding
 CC additional GalNAc moieties. Some peptide transferase activity is
 CC however not excluded, considering that its appropriate peptide
 CC substrate may remain unidentified. Prefers the diglycosylated
 CC NUC5AC-3/13 as substrate.
 CC -1- COFACTOR: Manganese and calcium (By similarity).
 CC -1- PATHWAY: Glycosylation.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By
 CC similarity).
 CC -1- TISSUE SPECIFICITY: Expressed in developing oocytes and egg
 CC chambers.
 CC -1- DEVELOPMENTAL STAGE: Expressed during embryonic, larval, pupal and
 CC adult stages. Weakly expressed during early embryonic stages but

CC displays a dramatic increase at 12-24 h of embryonic development.
 CC Continues to be in adult but displays much lower levels in the
 CC female adult as compared with the male.
 CC -1- DOMAIN: There are two conserved domains in the glycosyltransferase
 CC region: the N-terminal domain (domain A, also called G1 motif),
 CC which is probably involved in manganese coordination and substrate
 CC binding and the C-terminal domain (domain B, also called
 CC gal/GalNAc-T motif), which is probably involved in catalytic
 CC reaction and UDP-Gal binding (By similarity).
 CC -1- DOMAIN: The ricin B-type lectin domain binds to GalNAc and
 CC contributes to the glycopeptide specificity (By similarity).
 CC -1- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T
 CC subfamily.
 CC -1- SIMILARITY: Contains 1 ricin B-type lectin domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-eb.ch/announce/>
 CC or send an email to license@isb-eb.ch).
 CC -----
 CC DR EMBL; AY268065; AAQ56701.1; -;
 CC DR EMBL; AE003579; AAN10370.1; -;
 CC DR EMBL; AF24752; AAN7571.1; -;
 CC DR PDBase; FBgn0051956; Pgnat4.
 CC DR InterPro; IPR001173; Glyco_trans_2.
 CC DR InterPro; IPR008997; RicinB-like.
 CC DR InterPro; IPR000772; Ricin_B-lectin.
 CC DR Pfam; PF00535; Glycos_transf_2; 1.
 CC DR Pfam; PF00652; Ricin_B-lectin; 3.
 CC DR POSITE; PSS0231; Ricin_B-lectin; 1.
 CC DR Calcium; Glycosyltransferase; Golgi stack; Lectin; Manganese;
 CC Signal-anchor; Transferrase; Transmembrane.
 CC KW SIGNAL-ANCHOR; TRANSFERASE; TRANSMEMBRANE.
 CC FT DOMAIN 1 25
 CC FT TRANSMEM 26 48
 CC FT FT
 CC FT DOMAIN 49 659
 CC FT FT 192 303
 CC FT DOMAIN 360 422
 CC FT FT 511 644
 CC FT DISULFID 524 541
 CC FT DISULFID 571 588
 CC FT FT 615 632
 CC FT CARBOHYD 172 172
 CC FT CARBOHYD 194 194
 CC FT CARBOHYD 544 544
 CC FT CARBOHYD 580 580
 CC FT CARBOHYD 647 647
 CC FT FT
 CC SQ SEQUENCE 659 AA; 75805 MW; 0182D42AEBAA1DD2 CRC64;
 CC
 CC Query Match 36.2%; Score 1186; DB 1; Length 659;
 CC Best Local Similarity 41.2%; Pred. No. 2.9e-87;
 CC Matches 266; Conservative 106; Mismatches 216; Indels 58; Gaps 21;
 CC
 CC 2 RRKEKRLQAVLVLAALVLTLPVNGLMALYR-----ERQPDG---TPGSGAAYA 48
 CC DB 20 KRYVRLKRLKRVLLVLTIVTVSLVTLVVERRMKNAALTEQDLPDNGDITPFRANIH 79
 CC 49 PA--AGGSHSRKQKTFFLGDGQKL-----KDWMDKEAIRDAQVNGEGQR 94
 CC DB 80 PTKAKRPPFODNSVVDIPRSDKLGFRLEPBGKRGKMDHYAAMEADKKRGFGFEGHV 139
 CC 95 PYPM--TDAERV-DQAYRENGFIYVSDKISLRSLPDIRHPCNSKRYLETLPNTSII 151
 CC DB 140 AVKIENDEKQLEKEHYENMGFNGLSIRISVNRSVFDIRLEACKTRKYLAKLPNTSVIF 199
 CC 152 PFNNEGSSLLRTVSHVSLNRSPELVAAIYLVDDFSDREHLKKPLDYMAL-FPS-VRI 209
 CC DB 200 IFNENHNTLKRSTYSVNRIPBELKQIVLVDDGSEWDVLPKOPLDYVQGHFPHLVTV 259
 CC 210 RTKKRBLRTRMIGASVATGDIYTFLDSHCEANVMVLPPLDRIAANRKTIVCPMIDVI 269

DB 260 RNPEROGLIGARIGAKVAVGVGVWFEDSHIEVNYVMLPPLPIAINKISTCPWVDIT 319
 QY 270 DHDPFPERYQADPAMNGADPMWYKYRIRIPPELOKADSDPESPVMAGLFAVDRKMF 329
 DB 320 SHEDSFYFGNKDGAAGGDMKLYKQPLVLPB-DALDSMPFRSPVMNGGLPAINTDF 378
 QY 330 WEIGYDPLGLIWCGEQYEISFRKVMCGRMEDIPCSRGHLYR-KYVYKYPAGVS-LA 387
 DB 379 WDLGQYDDQDLIDGSGQYELSLFIMWCGMLLDVPSRSVAHI FRGPMKRGNDRGHNFYA 438
 QY 388 RNLKRYAEVWMDYAEIYQRRP-YRHLSDGVAVQKLRSLNCKSKRMFTKLAAML 446
 DB 439 KMKHRYAEVWMDYKQYVKKDPKYDNDAGDLTQRQVREHLKCKSKHMFWEVAVDPF 498
 QY 447 PKYVPEVPPAAMGIRIVVGTCL-CADTKHGLSLPLLEGVNRGRGAANNMGVTF 505
 DB 499 LVKFPVEPPSYAAGIIQVNAVPPVCLDMGKSTBEAVMFCADNRTPQPV-QFWEL 556
 QY 506 TWREDIRPDGPQHTKKF---CFDAISHTSP---VLYDCHSMKGNQLMKY-RKDKTLXH 557
 DB 557 STRFDR-----MKGFDSVCLDV--HEGPPNATVMMWCHGQGNQFPYVYRQTRLVH 608
 QY 558 PVSGS-CMD--GSESDHRIFMNTCNPSLITQWLFETHNSVLEKF 600
 DB 609 GERNKRCLEGFVENGIAKVAVNASCENGNDQRWFEFGVHTMDLTF 654
 CC
 CC RESULT 13
 CC AAQ56701 PRELIMINARY; PRT; 659 AA.
 CC AC AAQ56701;
 CC DT 02-MAR-2004 (T-EMBLrel. 27, Created)
 CC DT 02-MAR-2004 (T-EMBLrel. 27, Last sequence update)
 CC DT 02-MAR-2004 (T-EMBLrel. 27, Last annotation update)
 CC DE UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase (EC
 CC 2.4.1.41).
 CC GN Drosophila melanogaster (Fruit fly).
 CC OS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 CC OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC OC Ephydroidea; Drosophilidae; Drosophila.
 CC OX NCBI_TaxID=7227;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX PubMed=12829714;
 CC RA Ten Hagen K.G., Tran D.T., Gerken T.A., Stein D.S., Zhang Z.;
 CC RT "Functional Characterization and Expression Analysis of Members of the
 CC RT UDP-GalNAc:Polypeptide N-AcetylGalactosaminyltransferase Family from
 CC RT Drosophila melanogaster";
 CC RJ J. Biol. Chem. 278:35039-35048(2003).
 CC DR EMBL; AY268065; AAQ56701.1; -;
 CC KW Glycosyltransferase; Transferrase.
 CC SQ SEQUENCE 659 AA; 75805 MW; 0182D42AEBAA1DD2 CRC64;
 CC
 CC Query Match 36.2%; Score 1186; DB 2; Length 659;
 CC Best Local Similarity 41.2%; Pred. No. 2.9e-87;
 CC Matches 266; Conservative 106; Mismatches 216; Indels 58; Gaps 21;
 CC
 CC 2 RRKEKRLQAVLVLAALVLTLPVNGLMALYR-----ERQPDG---TPGSGAAYA 48
 CC DB 20 KRYVRLKRLKRVLLVLTIVTVSLVTLVVERRMKNAALTEQDLPDNGDITPFRANIH 79
 CC 49 PA--AGGSHSRKQKTFFLGDGQKL-----KDWMDKEAIRDAQVNGEGQR 94
 CC DB 80 PTKAKRPPFODNSVVDIPRSDKLGFRLEPBGKRGKMDHYAAMEADKKRGFGFEGHV 139
 CC 95 PYPM--TDAERV-DQAYRENGFIYVSDKISLRSLPDIRHPCNSKRYLETLPNTSII 151
 CC DB 140 AVKIENDEKQLEKEHYENMGFNGLSIRISVNRSVFDIRLEACKTRKYLAKLPNTSVIF 199
 CC 152 PFNNEGSSLLRTVSHVSLNRSPELVAAIYLVDDFSDREHLKKPLDYMAL-FPS-VRI 209

Db 200 IFNEHNTLLRSYIVNRPPELLKQIVLVDSGEWDVYKQPLDDYVQGHFPHLTVI 259
 Qy 210 RTKRGGLRTMGLASVATGVITPLDSCGANWMLPPLLRIRANRKTIVCPMDIV 269
 Db 260 RNBEROGLIGARIGAKVAVGVWVFDSDHIEVYVWMLPPLIRIAPKISCTCPMDVI 319
 Qy 270 DHDFPFYETAGAMGAFPMWYVYKRIPIPELQKADPDDPESPMAAGLFAVDRKMF 329
 Db 320 SHEDSFYSGNKGARCGFPMKLYKQLPVLPD-DALDKSMPIKSPFPMWGLFALINDFF 378
 Qy 330 WEIGYDPCGLIENGGEQYEISFKVMGGMEDIPCSRGVHYR-KYVPYVAGVS-LA 387
 Db 379 WDLGVDQDQIDWGEQYEISFKIMWGGLLVPCSRVAHITRGPKRPNRGHNFVA 438
 Qy 388 RNLKRAVAVWMDYAEIYQRPB-YRHSAGDVAVOKLRSSLNCKSPFMTKIAMWL 446
 Db 439 KKHKRAVAVWMDYAEIYQRPB-YRHSAGDVAVOKLRSSLNCKSPFMTKIAMWL 498
 Qy 447 PKFYPPVPPAAMGELRNVTGTL-CADTRHGALGSPLEEGCVRGRGEAAMNMVFTF 505
 Db 499 LVKPPVPPVYAGIIONVAMPYCLDNNGKSTEEAVGMSCADNTHPQPN--QFWE 556
 Qy 506 TWREDIRPGDPQHTKPF---CPDAISHTSP---VTLYDCHSMKGNQIMTKY-RKDKTLVH 557
 Db 557 SIFRDLR-----MKGFDSVCLDV--HEGPNATVWVMSCHSGQGNQFWYTDQOTQLVH 608
 Qy 558 PVSGS-CMD--CSBSHRIRFMTNCPSSLTQOQLFEHTNSTVLEKF 600
 Db 609 GENNKCLBEGFVNGIAKVAVANSCNDNRQWFEFVHNTMLDTF 654
 RESULT 14
 GLT9_DROME STANDARD; PRT; 650 AA.
 ID GLT9_DROME
 AC QMRG9; Q9V7T0;
 DT 01-OCT-2004 (Rel. 45, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Putative polypeptide N-acetylglucosaminyltransferase 9 (EC 2.4.1.41)
 DE (Protein-UDP acetylglucosaminyltransferase 9) (UDP-
 GalNAc:polypeptide N-acetylglucosaminyltransferase 9) (pp-GalNAc
 9).
 GN Name=pgant9; ORFNames=CG30463; fly).
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxId=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RC MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 Ramanathan P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle J.D.,
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 Bales K.M., Baas A., Baxendale J., Bayraktoglu L., Beasley E.M.,
 Beeson K.Y., Berens P.V., Berman B.P., Bhandari D., Bolshakov S.,
 Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 Burdette K.C., Busam D.A., Butler H., Cadenhead L.B., Davies P.,
 Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davis P.,
 De Paulis B., Delcher A., Deng Z., Dey A.D., Dew I., Dietz S.M.,
 Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 Durbin K.J., Evangelista C.C., Ferraz C., Ferrieres S., Fleischmann W.,
 Foerster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 Glodde C., Gong F., Gorrell J.H., Gu Z., Guan P., Harris K.,
 Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 Jaiswal M., Kalish F., Karpen G.H., Ke Z., Kienison J.A., Ketchum K.A.,
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Laske P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 Merklov G., Milshina N.V., Moberly C., Morris J., Moshrefi A.,
 Mount S.M., Moy M., Murphy B., Murphy L., Murthy D.M., Nelson D.L.,
 Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclet J.M.,
 Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 Svirskas R., Tector C., Turner R., Venter B., Wang A.H., Wang X.,
 Wang Z.-Y., Waasman D.A., Weinstein G.M., Weisenbach J.,
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zhang L.,
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195 (2000).
 RN [2]
 RP REVISIONS.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Mista S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 Hradecky P., Huang Y., Kaminke J.S., Milburn G.H., Prochuk S.E.,
 Smith C.D., Tupy J.L., Whitfield E.J., Bayraktoglu L., Berman B.P.,
 Bettecourt B.R., Celisner S.E., de Grey A.D.N.J., Drysdale R.A.,
 Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,
 Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 Lewis S.B.;
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
 systematic review."
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22 (2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley; TISSUE=Embryo;
 RX MEDLINE=22426066; PubMed=12537569;
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
 George R.A., Garin H., Krommiller B., Paclet J.M., Park S., Wan K.H.,
 Rubin G.M., Celisner S.E.;
 RT "A *Drosophila* full-length cDNA resource."
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8 (2002).
 CC -1- FUNCTION: May catalyze the initial reaction in O-linked
 galactosamine biosynthesis, the transfer of an N-acetyl-D-
 galactosamine residue to a serine or threonine residue on the
 protein receptor (By similarity).
 CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =
 UDP + N-acetyl-D-galactosaminyl-polypeptide.
 CC -1- COFACTOR: Manganese and calcium (By similarity).
 CC -1- PATHWAY: Glycosylation.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By
 similarity).
 CC -1- DOMAIN: There are two conserved domains in the glycosyltransferase
 region: the N-terminal domain (domain A, also called G1 motif),
 which is probably involved in manganese coordination and substrate
 binding and the C-terminal domain (domain B, also called
 Gal/GalNAc-T motif), which is probably involved in catalytic
 reaction and UDP-gal binding (By similarity).
 CC -1- DOMAIN: The ricin B-type lectin domain binds to GalNAc and
 contributes to the glycopeptide specificity (By similarity).
 CC -1- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T
 subfamily.
 CC -1- SIMILARITY: Contains 1 ricin B-type lectin domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AEO03806; AAF57964.2; -
 CC EMBL; AY121661; AAF51988.1; -
 CC HSP; P26514; 1KRM.
 DR FlyBase; FBgn0050463; CG30463.

DR Interpro; IPR001173; Glyco trans 2.
DR Interpro; IPR009061; Putativ DNA bind.
DR Interpro; IPR008997; Ricin like.
DR Interpro; IPR000772; Ricin B lectin.
DR Pfam; PF00535; Glycos transf_2; 1.
DR Pfam; PF00652; Ricin_B_lectin; 3.
DR SMART; SM00458; RICIN; 1.
DR PROSITE; PS0231; RICIN_B_LECTIN; 1.
DR Calicum; Glycosyltransferase; Golgi stack; Lectin; Manganese;
KW Signal-anchor; Transferase; Transmembrane.
FT DOMAIN 1 11
FT TRANSMEM 12 31
FT 32 650
FT DOMAIN 208 317
FT 317 590
FT DOMAIN 378 440
FT 440 643
FT DOMAIN 521 643
FT 643 90
FT 90 554
FT DISULFID 535 554
FT 554 577
FT DISULFID 577 590
FT 590 631
FT CARBOHYD 321 321
FT 321 321
FT CARBOHYD 373 373
FT 373 373
FT CONFLICT 145 145
FT 145 145
FT CONFLICT 454 454
FT 454 454
SQ SEQUENCE 650 AA; 73192 MW; AC847736AD1C07CA CRC64;

Query Match 35.8%; Score 1172.5; DB 1; Length 650;
Best Local Similarity 40.2%; Pred. No. 3.6e-86;
Matches 272; Conservative 91; Mismatches 201; Indels 113; Gaps 22;

2 RRKEKLLQAVLVLAALVLPVNGALYERQPDTPGSGAAVAPAGGSG-----55
6 RRSSTIVKLVLAALVLAALVLPVNGALYERQPDTPGSGAAVAPAGGSGAGLGP 65
56 -----HSRQKTFPL-----GDGQKXDMND-----76
66 IALALNRPAGEDEFGINGNVIGGGQ--KQAHDEADIPTVGKHKADLQAEHRKKAAB 123
77 -KEAIRDAQV-----GNGEGRPY-----PMTD--AEKVDAQYRNGNIVYSDK 120
124 PKKKPEDSKKVLDPRANFEENPGEIGKPVRLPKKMSDEMKKAVDDGWTNAENQVYSD 183
121 ISLNSRLPDIRHNCSK--RYLETLPYTSIIIPHNESGLRTYHSVYLRSPPELVAE 179
184 ISVHRLPDRDWMCDDEARYLNLPRKTVIICFHNEMVTLRTYHSVYLRSPPELVAE 243
180 IYLVDFSDREHLKKPLEDYMALFPYVRLRTYKREGLIRTRMLGASVATGDIYFLDSH 239
244 IYLVDFSDREHLKKPLEDYMALFPYVRLRTYKREGLIRTRMLGASVATGDIYFLDSH 303
240 CEANVMVLPRLDIRANRKTIVCPMIDVIDHDHDFRE--TQAGDARAGADMEYKRLP 298
304 CECTEGMLELRLIRANSTTVVCPVIDVSDTLEHYRDSGGVNVGDFMVLQFSWHP 363
299 IPELQKADPS--DPESPVMAGGLFAVDKFMELGQYDPGAEIINGEYOYSIFKVMK 356
364 VPEREKRNHSTLEPYSPFMAGGLSIDREPFDRIGTYSQYVYHIGNDKDG 423
357 GGMEDIPCSRGVHRYKRVYVPAVS--LAENLRVAEVMDEYAEYIYOR-----RPE 411
424 GGLTLEIVPSCGHVHIFRKSGPYKMRSGVNVLLKKNSVRLAEVMDVSEYVYHIGNDKDG 483
412 YRLLSGDAVAVQKCLASSLNCKSFKFMTKIAMDLPKTFYPPVPP--AAAGETIRNVG 469
484 W-----GDVSDRKLNDLCKSKFYLDNI-----YBELFIPGDSVNHGIRNLGYG 531
470 --LCADTKG-----ALGSPRLLEGCVRGGEAANNMNVFTFTWRDIPRDPQHTK 521
532 GRCTCLDPAKKGKQKAVGT-----YPCHRGGNQY-----MLSKAGETIRNDS-----576
522 FCPDAISHTSPVTLVDCHSMKGNQVWKYRQD--KTLVHPVSGSCMDSSESDHRIFNNTCPN 580

DB 577 -CIDYAG--KDYTLFECCHGKGNQFWTYRENTKQHLHGTSGKCLAISSKDXLMRECSA 633
QY 581 SSLTQOVLFEHNTSTVL 597
DB 634 SUSRQOVTLENYDSKL 650

RESULT 15

Q70815 PRELIMINARY; PRT; 645 AA.

ID Q70815
AC Q70815;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE BLIP5044 (Fragment).
GN Name=eb1G5044; ORFNames=EN5ANG0000003900;
OS Anopheles gambiae str. FEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=PEST;
RC Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is preliminary data.
DR EMBL; AA01008944; EMBL0180.1; -.
DR Interpro; IPR001173; Glyco trans 2.
DR Interpro; IPR000772; Ricin B_lectin.
DR Pfam; PF00535; Glycos transf_2; 1.
DR Pfam; PF00652; Ricin_B_lectin; 3.
DR PROSITE; PS0231; RICIN_B_LECTIN; 1.
DR NON_TER 645
SQ SEQUENCE 645 AA; 73459 MW; A4DFD6CE16B51C0 CRC64;

Query Match 35.4%; Score 1159; DB 2; Length 645;
Best Local Similarity 44.3%; Pred. No. 4.4e-85;
Matches 251; Conservative 79; Mismatches 193; Indels 44; Gaps 19;
QY 38 GTPGSGAAVAPAGGSGSHRQKTFPLDGGKLMKMDKEAIRDAQVGN--GEGRP- 95
101 GDRGGYGGC---CGGRDADSSMPRTY---RPQELKKMRQAPTV--AEKVYRGEGWKPV 151
DB 96 -YPMIDAEVDAQYRNGNIVYSDKISLNSRLPDIRHNCSKRYLETLPNTSIIIPH 154
152 KIPANQOELMKSKFKENQFNLSADMTWLNRSLTJVRHNDCKKQYPAKLPTTSIVYH 211
QY 155 NRGMSLRTYHSVYLRSPPELVAEIVLVDFSDREHLKKPLEDYMALFP--SVRLRTYK 213
212 NEAMSTLTTISVYLRSPPELVAEIVLVDFSDREHLKKPLEDYMALFP--SVRLRTYK 271
214 REGILRTRMLGASVATGDIYFLDSHCEANVMVLPRLDIRANRKTIVCPMIDVIDHD 273
272 RGLIRARLLGAKHVAGQYITFLDACECTEGMLELRLIRYLDRTKTVCPIIDVIDST 331
QY 274 FYETQAGAMGAFPEWYRYKRIPLP--ELQKA--DPEDPESPVMAGGLFAVDKFMK 330
332 FEVYLT--ASQUTGGFWMKLFPMYRVPAEMQRHNDRTAPLRTPMAGGLSIDRDYV 390
331 ELGGYDPGAEIINGEYOYSIFKVMKCGMEDIPCSRGVHRYKRVYVPAVS--LAR 388
391 EIGSIDEGNDINGENLESPFVMMCGGLEIAPCSRGVHRYKRVYVPAVS--LAR 450
DB 389 NLKRVAEVMDEYAEYIYORREYHLSAGDAVAVQKCLASSLNCKSFKFMTKIAMDLPK 448
451 NNAARLAEVLDMSSEYVYININPARKASAGDVSEBRALAEKSKSFYRLDNI-----504
QY 449 FYPPVPPPAAM--GEIRNVGTGLCADYKTHGALGSPRLLEGCVRGGEAANNMNVFTT 506
505 -YVESQMPLDYFLGEIRNVKTHNCIDTIGRKSNEKIGSSYC--HGIG-----GNQVFAVT 557

Qy 507 WRSDIRGDPQHTKKFCFEDAISHTSPVTLYDCHSMKGNOLWKY-RKDKTLYHPVSGSCMD 565
Db 558 KRHQIMSDN-----CLDASNALGPNLVRCBGMGNGNEMIIDDEKTIKAVNSGNCILT 611
Qy 566 -CSESDHRI-FMNTCNPPSLTQOMLFE 590
Db 612 RASEDDPSTPLRRCNYSB-GQOWLMQ 637

Search completed: December 20, 2004, 14:12:57
Job time : 200 secs

This Page Blank (uspto)